

When will the Gulf War be over?

The shooting may have stopped on the border between Iraq and Kuwait, but the prospects for lasting stability in the Middle East seem remote. The region needs a demonstration that hopeful developments are possible.

THE shooting part of the Gulf War is over, but its end is not yet in sight. Rarely can there have been such a quick war whose aftermath has been so prolonged. The difficulty is banal; although the war has formally reversed its own immediate cause — Iraq's annexation of Kuwait — the present differs remarkably from the *status quo ante bellum*. It is as if a quantum measurement had been carried out in the Middle East, leaving the system in a state that differs substantially from its initial state.

The most obvious difference is that the war is not yet formally at an end. That cannot happen until the United Nations Security Council has resolved on a way of unwinding the events of the past eight months that offers some prospect of stability on the border between Kuwait and Iraq, and perhaps for Iraq itself. That would be easier if the present government of Iraq were not battling for its own survival against the wishes of substantial numbers of its people. The draft resolution on which the permanent members of the Security Council are working would establish a lien on Iraqi oil revenues and maintain sanctions against Iraq until its chemical weapons and ballistic missiles are destroyed.

The present intention is that these conditions should be externally devised and then presented as a *fait accompli* to Iraq; it would be far preferable if they could be negotiated with a government willing to acknowledge that the invasion of Kuwait was wrong; then, for example, it would be possible to arrange that the terms of reparations are not dangerously over-onerous. Meanwhile, nobody should be surprised if troops from the Gulf coalition (chiefly US troops) are still in southern Iraq long after this time next year.

Kuwait itself will not easily be returned to its former state, which may even be unattainable. Not only are there 600 burning oil wells to extinguish, and rudimentary utility services still to be restored, but the Iraqi occupation has understandably scarred the tiny community of Kuwaitis (fewer than 700,000). It is understandable that the remnants of the most sophisticated communities in the Gulf should now be muttering about the need for a measure of democracy.

Hostility

It is less encouraging that the same community is reported to have reacted with sometimes unconcealed hostility towards non-Kuwaiti Arabs, Palestinians especially. When the oil-well fires are put out, Kuwait will again discover that 600,000 people on their own cannot manage one of the world's biggest oil-production industries. The academic community

elsewhere should help Kuwait's two reputable institutions, the university and the research institute (see page 369).

In Kuwait, as elsewhere in the arabic and predominantly Moslem oil-rich Gulf, questions of equity are stumbling-blocks to change, partly because they are hardly ever raised. Governments with feudal (but often recent) origins have slowly been persuaded to modernize their autocracies, but to a variable and inadequate degree. In most places, autocracies have mostly become more benevolent, the more easily as oil revenues have increased. Parliaments of a kind have cropped up only in the smaller states, Dubai and the United Arab Emirates, for example. (That in Kuwait was suspended in 1986, but may now be restored.)

Prosperity

But to the simple question "Who owns the oil?", there is no simple answer; Gulf nationals enjoy the prosperity paternally shared out, appointed (not elected) governments collect the proceeds of selling oil, but hardly anybody knows who owns the surpluses left over. There is no obvious reason why cupiditous curiosity on questions such as that should be suppressed only among readers of the Koran.

The social problems of the oil-rich Gulf would be more tractable if the indigenous population were greater. The population of Saudi Arabia, physically three-quarters of the Middle East, is, at 7 million, less than half that of Iraq. The population of all the Arab oil-rich states is less than half that of Farsi (and Moslem) Iran, the other significant oil-producer in the region whose populations is now approaching 60 million. The result is that the Arab oil-producers depend on armies of itinerant workers for the production of their wealth and even (by means of service industries) for assistance in its enjoyment.

This practice has indirectly won economic benefits for countries as far apart and as different as Egypt and the Philippines, whose citizens have been among the most pathetic of the refugees from the Middle East since last August. The same practice has helped to blunt the edge of social change in the oil-producing states; however well paid, migrant workers do not qualify for citizenship, which example saps the confidence of paid-up citizens seeking equity in the conduct of their own affairs. Paradoxically, stability in the Middle East might be assisted if the Organization of Petroleum Exporting Countries (OPEC) were one day to realize its old dream of reducing production quotas in the Middle East, but there seems very little hope of that.



Nor is there much prospect of a softening of the conflict between Israel and its neighbours. It is curious that the United States, which took the initiative last August in defending Saudi Arabia from the threat of an attack by Iraq, did not demand as part of the price for its intervention an undertaking that the security guarantees that Israel needs would eventually be forthcoming. Mr James Baker, the US Secretary of State, seems equally to have been disappointed on his recent swing through the Middle East, as also in persuading the government of Israel to trade territory (occupied illegally) for peace. It is to be hoped that President George Bush will have better luck when he goes to Jerusalem this month.

Security

Israel has a right to expect that its security for the indefinite future should be assured, but it cannot continue rejecting all proposals for a humane and seemingly settlement of the Palestinian problem, and not making realistically constructive suggestions of its own, without running serious risks. It is a shame that the most sophisticated, and technically advanced, government in the region should be able to act only as if it were slave to events created by others.

The Gulf War has also cast a shadow well outside the Middle East. No doubt, for example, there are generals in the Soviet Union who have been made to think twice by the demonstration in the past three months of the destructiveness of even conventional warfare. On the face of things, that should make it easier for the Soviet Union to comply with the letter of the treaty on conventional arms in Europe, signed with such a flourish less than a year ago, but still to be ratified.

But the argument can also run the other way. If the generation of Western munitions planned 15 years ago are as effective as they are advertised (by their users) to have been in the battle against Iraq, can it be wise to scale down the Soviet forces deployed west of the Urals to the degree required by the treaty? Mr Mikhail Gorbachev may find himself compelled to give comfort to his military in this respect in exchange for their assistance on the streets of Soviet cities in the months ahead. That would be bad news.

Intrigue

So is the discovery of the sophistication of the network of intrigue by which Iraq has been supplied, over several years, with military technology from Western Europe and North America. There are elements of black comedy in the business — last year's hare-brained scheme for supplying Iraq with components of a monster gun for example, the discovery by US troops in the Gulf that their adversaries were occasionally equipped with night-vision equipment manufactured in Texas and the revelation that a branch office in Atlanta of an Italian bank had made \$3,000 million of credit available for Iraqi purchases of equipment.

The reprehensible ambivalence of Western governments, especially during Iraq's war with Iran, towards the sophistication of Iraq's military development may have spread confusion needlessly, but the evident willingness of companies in Western Europe and even the United States to circumvent

the law on military exports in the search for apparent profit is a sobering lesson. The defence committee of the British House of Commons is embarking on an inquiry into recent British practice, as if to learn how many horses bolted through the stable door. The irony is that the military in the West, zealously efficient at keeping secrets from their own civil industries, seem powerless at preventing their loss to irresponsible governments elsewhere.

Sadly, the legitimate trade in arms seems destined to continue. Almost before the shooting in Iraq and Kuwait had ended, the United States had made arrangements to provide Egypt and Saudi Arabia with advanced military equipment. There seems to be general agreement with the proposition that the governments of the Middle East are too well-armed, but no understanding about a mechanism by which that state of affairs might be abated. The Canadian government has been pushing for an international agreement on the arms trade in the Middle East, but the best that can be hoped for is a mechanism by which arms-supplying governments can tell which of their customers has what equipment on what scale. In the long run, only an agreement between the states of the Middle East itself, Israel included, can bring this ruinously expensive and dangerous trade within seemingly limits.

Conflict

So is there nothing that can be done without waiting on a settlement of the most obdurate conflict of all, that between Israel and its neighbours? That is the chief log-jam in the Middle East, as it has been for many years, but not everything need wait on its resolution. Egypt, which has a peace treaty with Israel and which is a crucial member of the Gulf coalition, has an important part to play. So too does Jordan, which lost too many friends for its own comfort during the Gulf War, which is the *de facto* home of a great many Palestinians and whose economic prospects are among the poorest in the Middle East (especially now that rich Saudi Arabia has taken offence at its support for Iraq). So, strange as it may seem, does Israel. And so do the rich countries of the West.

The most readily forgotten consequence of the war is the uncounted toll of death in Iraq. Next to that, and less easily overlooked, is the misery caused by the displacement of people in the Middle East, many of whom have finished up in Jordan because they have nowhere else to go. It would help to bring civility to the Middle East if those who wish it were allowed (with Jordan's consent) to settle where they are, and if those who dream of a New World Order were to embark on demonstrating how that community could be made at once economically viable and socially coherent. Israel, after all, has the technical competence to make the desert bloom, and every incentive to begin to deliver on its implied promise that it would provide that service for the whole Middle East if only there were peace. Jordan would have nothing to lose from such a project, but would gain the assurance that the refugees will not further drain its economic resources. And Egypt is well-placed to be peacekeeper. The truth is that after a miserable eight months, the whole Middle East needs practical proof that resources and ingenuity can conjure civility and a sense of progress out of seeming nothing. □

US plans to rescue materials research

- Government-wide effort to boost funding
- Would resemble supercomputer initiative

Cincinnati, Ohio

THE White House Office of Science and Technology Policy (OSTP) is planning a major initiative in materials research, to rescue a hot field threatened by an ice bath of inadequate funding.

The initiative is expected to include "hundreds of millions in new and reprogrammed dollars" for materials research, said David Huber, executive director of the OSTP-affiliated White House National Materials Council, speaking at the American Physical Society's annual meeting in Cincinnati late last month. The proposal is scheduled to be included in the 1993 White House budget request to Congress.

Huber's announcement brought applause from the researchers in the audience, many of whom have seen their funding opportunities wane even as their discoveries in high-temperature superconductivity, semiconductor physics and other areas of materials science have drawn praise for both their scientific and commercial value. Although the number of researchers in the field has doubled in the past decade, funding has remained essentially level, rising only as fast as inflation. Academic scientists have been forced to turn almost exclusively to their own departments to fund new laboratories, and young investigators now actually have a better chance of winning a special prize or fellowship than a regular grant from the National Science Foundation (NSF).

The fact that twice as many scientists could scrimp and save to survive on the same amount of money, "proves that physicists are very compressible, but it doesn't bode well for the future", says Michael Schluter, an AT&T Bell Laboratories physicist who headed a recent study¹ on the subject by the American Physical Society.

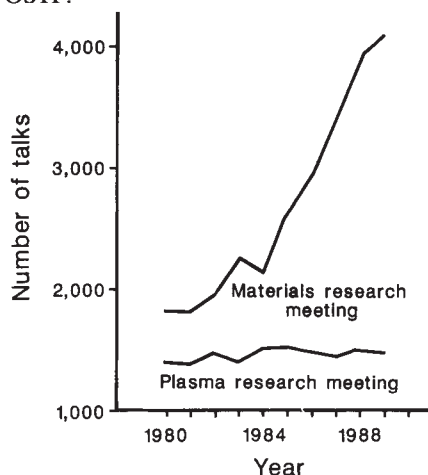
Describing the field as "seriously threatened", the report finds that support for individual investigators has been especially hard hit. For example, NSF, the chief supporter of university research, is spending some \$100 million less than it was ten years ago.

"The times are dramatically out of joint for Condensed Matter Physics," the report warned. "In an era where this field has become one of the very most active and relevant of the sciences, its future is in jeopardy."

Earlier this year, materials researchers released another report² calling for \$1,250 million a year of new funding from industry and government (an increase of about a third over current spending) to remain competitive with the rest of the world. Huber told the physicists that OSTP is "paying a good deal

of attention" to that report and would hammer out the details of the new effort with input from the community.

OSTP intends to organize the initiative much like the new high-performance computing and education initiatives the White House announced in February, and a global warming research initiative created the year before. Like those, the materials initiative will be spread across all the US agencies that currently do some materials research, coordinated by an inter-agency body known as the Federal Coordinating Council on Science, Engineering and Technology (FCCSET, pronounced "fix-it"), which is headed by D. Allan Bromley, director of OSTP.



The number of material-research scientists, as measured by talks at the American Physical Society's March meeting on condensed-matter physics, has more than doubled in the last decade, while those in other disciplines (plasma physics, for example) showed little change.

Warning that "for a more complete story, you'll have to wait for the 1993 budget", Huber nevertheless described a planning process that is already well under way. OSTP is now taking an inventory of the \$1,600 million the federal government is already sponsoring in materials research, in order to identify overlaps or opportunities for collaboration. Rough plans for the initiative will be completed by mid-June, Huber said, and will then be circulated to the relevant agencies for comment. Final approval will be up to the powerful Office of Management and Budget (OMB), which will set the actual dollar request. **Christopher Anderson**

1. *Opportunities, Resources, and Problems in the Field of Condensed Matter Physics*, American Physical Society, 1990;
2. *A National Agenda in Materials Science and Engineering*, Materials Research Society, February 1991.

SCIENTIFIC MISCONDUCT

UK medical guidelines proposed

London

BRITISH research funding agencies should consider denying grants for medical research to institutions that have no mechanism in place to investigate allegations of scientific misconduct, according to a Royal College of Physicians (RCP) working party. The working party's report, published this month in the RCP's official journal, is the first attempt by the British medical research community to tackle scientific misconduct, and suggests a strict procedure for research institutions to follow when allegations arise.

The RCP working party was set up in January 1989, largely on the initiative of Stephen Lock, the now-retiring editor of the *British Medical Journal*. In 1988, Lock conducted a confidential survey of prominent British medical researchers and journal editors. He found that only three of the 29 research institutions canvassed had any formal mechanism in place to investigate allegations of misconduct. And more than half of the 79 individuals who responded knew of an instance of misconduct — suggesting that the problem in Britain is much wider than the half-dozen documented cases of fraud, plagiarism or piracy of ideas.

The RCP recommendations borrow heavily from the 1989 guidelines laid down by the US Institute of Medicine. Students should be introduced to a code of practice for research, young investigators' raw data should be scrutinized frequently by their supervisors, and all data should be retained for ten years, the working party says.

The RCP suggests a blueprint for research institutions to follow in handling fraud, modelled on the 1988 guidelines of the Association of American Universities. But the RCP recommendations fall short of calling for a central body to investigate allegations of misconduct, along the lines of the US National Institutes of Health's Office of Scientific Integrity. With no central body to apply pressure from above, research institutions may be inclined not to investigate allegations against their researchers, to avoid attracting negative publicity. But Tawfique Daneshmend, from the Royal Devon and Exeter Hospital, and a member of the RCP working party, says that the recommendations should be seen as the "first word" on misconduct from the medical research community — the question of a central investigating authority is not closed.

Although most of the RCP recommendations are uncontroversial, the suggestion that funding bodies should award grants only to institutions that have a mechanism to investigate misconduct seems likely to divide the research community.

Peter Aldhous

Good news, for a change

Munich

WHEN the German science advisory council Wissenschaftsrat began last year the monumental and delicate task of evaluating all the research institutes of the former East German Academy of Sciences, researchers in both parts of Germany were fearful that German science, once unified, would emerge as less than the sum of its parts. East Germans, in particular, feared jealousy and vindictiveness from the Western evaluators, who were seen as competing for pieces of the same finite pie. "They (eastern German scientists) thought all we wanted was to turn out the lights" in their institutes, recalls Dieter Simon, chairman of Wissenschaftsrat.

But as the evaluation approached its midpoint last week, those fears have been laid to rest. A scientific infrastructure is beginning to take shape, at least on paper, in which the whole may be even more than simple addition would have indicated.

There is good news, for instance, on how many researchers from several former East Germany Academy institutes will have positions in new institutions that are now being

planned. Two weeks ago, Heinz Riesenhuber, the German Research and Technology Minister, estimated that 7,000–10,000 scientific personnel could be kept on out of approximately 15,000. The 5,000–8,000 lost positions, while unfortunate, are less than many had feared.

And in some areas, Wissenschaftsrat would like to retain an even higher percentage of researchers. In information sciences, for instance, the council recommends keeping as many as two thirds of researchers who have not already been hired by industry. In some mathematics institutes, the percentage may be even higher.

As he has done in earlier recommendations, Simon warns that the governments in Bonn and the *Länder* (states) must provide adequate funds to support the new institutions. Simon, who is known for not being easily ruffled, adds that the possibility that the old institutes could be shut down while their successor institutions are postponed for lack of funds, "gives me sleepless nights".

Under severe political pressure from the nearly bankrupt eastern *Länder*, Chancellor

Helmut Kohl agreed on 28 February that Bonn would give them an additional DM24,000 million. Although he sees the grant as an encouraging sign, Simon points out that the *Länder* have not yet promised to spend any of the new money on science. "It is still an open question whether anything filters down at all," he warns.

Perhaps the most promising sign for research in eastern Germany is the decision by Wissenschaftsrat to propose the establishment of the first Max Planck Institute in that part of the country. The proposal for a new institute in Halle dedicated to solid-state physics and electron microscopy is a recognition that the existing academy institute there is a "jewel of the former East Germany", as council spokesman Wilhelm Krull puts it.

But the recommendation is not without problems. For one, the Max Planck Gesellschaft, which would be expected to run the institute, already has highly respected institutes for solid-state physics (in Stuttgart) and electron microscopy (in Berlin). Nevertheless, Wissenschaftsrat was confident that there was little enough overlap to make a new Max Planck Institute feasible and desirable.

More challenging will be the need for

'Now it seems this whole enterprise is about to fail'

Munich

DIETER Simon, director of the Max Planck Institute for the History of European Law in Frankfurt, is chairman of the science council Wissenschaftsrat, which has been evaluating science in eastern Germany. Here, near the midpoint of the evaluation, he reflects on the emerging picture of how science in Germany is being unified.

"At the beginning, we all had good intentions. We saw the necessary unification of science and education as an opportunity to restructure our own obsolete system and that of East Germany and use our common resources to create a new structure. We were going to take clever advantage of a unique historical opportunity.

"Now it seems as if this enterprise is about to fail. What has happened? First, a miracle. Our much-criticized system of science and education in the West seemed to improve by the day as we learned more about the shortcomings in the system in the East.

"Sure, some of our professorships are not filled by the optimal people. But that problem pales by comparison to a system where membership in the Communist Party was a main criterion for selection.

"Sure, our universities had a poor student-faculty ratio, especially in these days of severe overcrowding. But over there, universities and institutes were not only immobile but also jam-packed full with unnecessary staff.

"Comparisons like these made it easy for us to forget our good intentions. On top of

that came practical problems. The reforms had to be carried out very quickly, so that students and professors alike did not have to wait around too long, their futures hanging in the balance, before a new structure was set up.

"But our long-term goal of basic structural reform in East and West and our short-term goal of giving concrete practical help to the East are working at cross-purposes.

"We simply do not have the time and distance to plan research centres that combine the advantages of non-university and university institutes and avoid their drawbacks. The same is true for facilities for practical training, for apprenticeships and for unconventional university courses. The Western system will simply be extended eastward.

"The next problem was a conflict of goals. It is not easy to bestow upon the East the benefit of our functioning system of doing science and at the same time for both sides to move forward toward a better common future. It has to go in stages: first compatibility, then reform.

"Thus the inherent absurdity of trying to bring back research in the East from non-university structures to universities, where, everyone in East and West agrees, it belongs. But in the West, just 45 per cent of federal research money goes to universities, a figure that shows no signs of changing. What kind of an example does this set?



Simon — good intentions.

Then there is the human factor. It is easy to make recommendations. But who puts all these changes into practice? Civil servants, of course. The same inflexible petty bureaucrats who choose, with a few notable exceptions, to stay in their offices in the West instead of going east where they are needed, leaving the ministries in the East helpless to implement our recommendations.

"New ideas seem always to come up short in such a system. The tried and true 'funding instruments' we know in the West are not applied because the current conditions do not fit rules that have been invented for other situations. Instead, research groups are shoved out the door and pointed toward a 'market' that does not exist. If we are not careful projects, and entire institutions will fall through the cracks because they do not fit the dimensions of a civil servant's desk.

"We still have good intentions. But that will not be enough. We need to increase our tolerance for unorthodox solutions. We need to go to the limits of what the two systems will tolerate in our attempt to unify them. We may have to live with institutions that make us uncomfortable.

"But we cannot forget that our system is not perfect either. We have to force ourselves to remember our shortcomings every day.

"In a lot of ways, we do know better. But not in every way." □

speed in assuring the institute's future before the best researchers are lured away. Although the Max Planck Gesellschaft has traditionally followed every recommendation from Wissenschaftsrat to found a new institute, it has sometimes taken years to open them, and then only when another institute has been closed or the Gesellschaft has received a budget increase.

A spokesman for the Gesellschaft, Michael Globig, says that an internal committee will be formed, probably in June, to examine the question. The committee is expected to make a recommendation by March 1992, about the possibility of taking over the Halle institute. But the recommendation, if it comes, can be carried out only if Bonn and the *Länder* make new money available. "We don't plan to close another institute to open this one," says Globig.

In another important development, Wissenschaftsrat is for the first time asking policy-makers to reduce the scope of western institutions in a particular field so as to shift part of the focus of that research to the east. The advisory council is urging the western German Aerospace Research Centre (DLR) and the eastern Institute for Cosmos Research (IKF) to work together in order to make the most of the experience of IKF. This could be best achieved, says the council, by founding a new DLR institute at IKF in Berlin for planetary remote sensing, and setting up two affiliated groups to be supported by the Max Planck Gesellschaft.

The recommendation is partly an explicit recognition of the excellence of the East German space flight programme, which was well funded because of its publicity and diplomatic value for the state, demonstrated in several joint space missions with the Soviet Union. And Wissenschaftsrat has determined that space science in eastern Germany is also outstanding, especially in remote sensing and planetary science — areas where, it notes, there are some gaps in the western research programme.

One of the most surprising results of the latest round of evaluations is the relatively strong support received from Wissenschaftsrat by researchers at former East German Academy institutes for economics and social sciences — which are generally considered to have been tainted by East Germany's state-enforced Communist ideology. Simon, himself a historian of law, asserts that evaluators found some outstanding groups of researchers, isolated even within their own institutes, who are worth saving. These researchers, Simon says, are less burdened by their past than he had thought, in part because their institutes had very little internal coherence and were divided into clusters of 5 to 8 people cut off from other groups.

Wissenschaftsrat recommends that a new institute for empirical economic research be set up in eastern Germany as well as a "commission for investigating the social and political transformation of the new *Länder*".

Steven Dickman

Where angels fear to tread...

Washington

FROM fetal tissue to animal rights, the new 102nd Congress is proving that, like its predecessor, it has not met a research controversy it did not like.

Last week, Representative Henry Waxman (Democrat, New York) fired a warning shot across the bow of the administration by introducing a bill that would reverse the contentious existing ban on federally funded fetal-tissue research. It would also make break-ins and protests at animal facilities a federal crime, and boost participation of women and minorities in federally funded clinical trials.

The bill is essentially identical to a proposal Waxman introduced last year, too late to be considered as the 101st Congress drew to a close (see *Nature* 348, 101; 8 November 1990).

Although the legislation — with the full force of the powerful US anti-abortion movement against it — is unlikely to pass as it stands, Waxman intends to use it as an opportunity to hold hearings at which he can grill administration officials on their fetal-tissue policy. Bernadine Healy, the new director of the National Institutes of Health, is expected to be called on 15 April to explain her defence of the ban (see *Nature* 350, 178; 21 March 1990).

Another provision in Waxman's bill joins other new and pending congressional legislation aimed at combating animal-rights attacks on research facilities. It would designate as federal crimes break-ins at federally funded health facilities, as well as protests that involve "obstruction through intimidation".

Federal offences carry stiffer penalties and

investigations than simple state crimes, as most animal-related break-ins are now considered.

In the Senate, a similar break-in bill was reintroduced earlier this year by Howell Heflin (Democrat, Alabama). Heflin's bill passed the Senate last year, but the 101st Congress closed before a matching bill could be passed in the House of Representatives. That companion bill — by Texas Democrat Charles Stenholm — is also to be reintroduced soon.

Stenholm is informally circulating his legislation to gain support, and is said to have more than 100 representatives ready to sign on as cosponsors. Although the three bills differ in some details, research lobbyists are focusing their energies on the Heflin and Stenholm package as the most likely to pass. On the opposite side animal advocates are also mobilizing their forces for battle, and a stiff lobbying war over the bills is expected.

On the theory that support is where you find it, the animal activists have established an uneasy alliance with the anti-abortion movement in opposition to the break-in bills. Last week, the National Right to Life Committee came out in opposition to the Waxman provision, arguing that activist protests at animal facilities are legally similar to protests at abortion facilities — something they would not like to see made a federal offence.

Using a slightly different rationale, the American Civil Liberties Union also opposes the bill, on the grounds that deterring legitimate protest violates the First Amendment right to free speech.

Christopher Anderson

LEGISLATION

A Bill of Rights for fruit and vegetables?

Washington

AN unusual piece of legislation in Colorado has caught the attention of vegetable hatters in that state. The Disparagement of Perishable Agricultural Food Products Act is aimed at discouraging slanderous campaigns about fruits, vegetables, meats and other produce.

The bill, which has passed both the Colorado State Senate and House of Representatives and needs only the signature of Governor Roy Romer to become law, has become the butt of many jokes in the press. Would President Bush be precluded from making his anti-broccoli sentiments public in the state of Colorado? Will the term 'meat-head' become actionable?

The bill is not as crazy as it sounds, says sponsor Steve Acquafresca, a legislator from a heavily agricultural part of the state. "There is a terrific need out there to provide victims of false food scares with the means

to recover all or part of their damages," says he. Not vegetables themselves, mind you, but food producers.

Acquafresca's bill was motivated in part by the Alar scare in the United States two years ago, when environmental groups launched a campaign to convince the public that small amounts of Alar, a preservative used on apples, would cause cancer. The resulting scare cost apple growers in Colorado and the rest of the country an estimated \$130 million. Under the new legislation, victims of food scares might be able to collect damages.

The requirements of the bill are fairly rigorous: in order to file a claim, the victim must show that whoever spread the rumour knew that the information was false, and it applies only to "malicious or negligent false food safety scares that are conveyed in a public campaign manner", says Acquafresca.

Diana Steele

Deforestation rate is falling

São Paulo

BRAZILIANS are burning their rain forest less. The latest figures on deforestation in Amazonia show that the felling and burning of trees diminished by 27 per cent from 1989 to 1990. The new data collected by the Institute of Space Research (INPE) refer to deforestation from August 1989 to August 1990 in the 4.6-million-square-kilometre region known as Legal Amazonia, which comprises parts of nine states in the north of Brazil.

In the 1989 study, INPE declared that 17,871 square kilometres of forest had disappeared. The new study puts the figure at 13,818 square kilometres. Data were provided by the Thematic Mapper instruments aboard the Landsat 5 satellite.

Previous studies by the Space Institute had a greater margin of error — 10 per cent — because the process of adding up the data was not automated. More advanced methods of processing the data made the margin of error fall to 5 per cent, according to the director of the institute, Márcio Nogueira Barbosa. But although in the past INPE has been forced to issue statements correcting data it had released, the biggest problems with some of the studies have been political. In the government of president José Sarney (1985–90), INPE tried to present the data in the most favourable way, 'making it up' to show small deforestation percentages. Sarney even declared to *Newsweek*, before INPE's studies

were released, that only 5 per cent of Amazonia was destroyed, when the actual figure is close to 10 per cent. INPE obediently said that the percentage was only 5 per cent, and later had to correct it and give further explanations.

President Fernando Collor de Mello tried to throw off the image of Brazil as environmental villain. The new figures, says José Goldemberg, the Secretary of Science and Technology, would make it easier for Brazil to deal with international institutions. He thinks that the smaller rate is a consequence both of the increase of fiscalization and the cut of subsidies for farming in Amazonia.

Even though the overall rate has decreased, there is cause for concern in some specific areas. Rondonia, the worst affected state in recent years, showed a slight increase in deforestation — 1,676 square kilometres against 1,441 in the previous study. The other state where deforestation increased, Amapá, has 9 per cent of its territory deforested, in Rondonia the figure is 14 per cent.

Brazilian government officials hope that this trend will continue, so that they could show the world the country's new pro-environment stance. The forum for this would be the United Nations Conference on the Environment and Development, scheduled to meet in Rio de Janeiro in 1992.

Ricardo Bonalume

From mine camp to laboratory

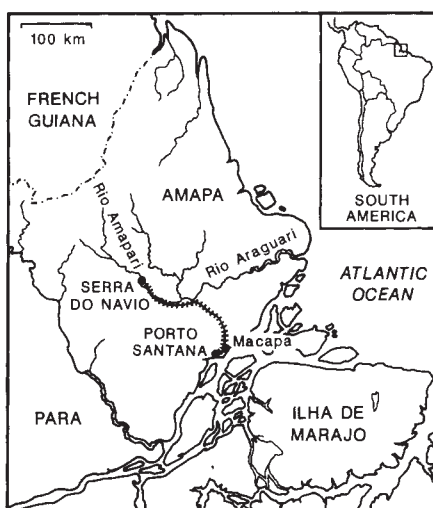
São Paulo

BRAZIL'S biggest university has received an unexpected gift: a new base for tropical research deep in the Amazonian rain forest, complete with facilities to house 2,000 people. The University of São Paulo (USP) announced last month that a Brazilian mining company is giving it a complete town, Vila Serra do Navio, in the state of Amapá, which borders French Guiana.

The company, Indústria e Comércio de Minério SA (ICOMI), obtained a government concession to mine manganese in Amapá in 1950. The concession expires in 50 years, but the manganese ran out first. So instead of simply abandoning the town for the jungle to reclaim, ICOMI offered it to the university, which gladly accepted.

The mining company said it was proud of having built a model town in the middle of the forest and did not want to see it simply disappear.

To maintain the town's infrastructure, the company spends about \$3 million a year, and ICOMI has promised to pay the university for the upkeep until the end of 1992. After that, the university must take over, but USP rector Roberto Leal Lobo e Silva Filho says it will not shift funds from the budget to pay for it. Brazil's Secretariat for Science and Technology has already



shown interest, he said, although it has not yet offered any money.

Vila Serra do Navio is 196 km by railroad from the Amapá coast. The town has about 500 houses, the only hospital in the area — which attends to the needs of about 12,000 local inhabitants — and shopping and recreational facilities. With the miners gone, the elementary school and several other buildings will be turned into laboratories.

Ricardo Bonalume

First maternal transmission?

London

A NEW case of bovine spongiform encephalopathy (BSE) may be the first known in which a cow was infected by its mother *in utero* rather than from infected feed.

Last week, pathologists from the Ministry of Agriculture Fisheries and Food (MAFF) confirmed the presence of BSE in a cow from Cheshire born in November 1988. Since the birth date was three months after the government banned the use of cattle feed containing ruminant meat and bone meal, it seems unlikely that the animal contracted BSE by that route. The cow's mother developed BSE in 1989.

John Wilesmith, who heads MAFF's BSE epidemiology team, says that his group must now examine the animal's history, to check that it was not fed accidentally with banned cattle feed. But even if confirmed, an isolated case of maternal transmission is not a cause for great alarm, Wilesmith says: maternal transmission would have to be a common occurrence significantly to lengthen the British BSE epidemic. MAFF has no plans to ban breeding from the offspring of BSE cases.

Peter Aldhous

COMPUTER INDUSTRY

ICL expelled from JESSI projects

London

THE UK-based computer company ICL has been expelled from three projects within the Joint European Submicron Silicon Initiative (JESSI), the \$5,000 million collaboration in semiconductor research funded by European computer companies, the European Communities (EC) and EC member governments. The decision comes in response to ICL's takeover last year by the Japanese company Fujitsu — a sensitive issue for JESSI, which was set up in 1989 to help the European computer industry to compete with Japanese companies.

The ejection from JESSI follows ICL's expulsion in January from the European Information Technology Round Table, a group of 12 European computer companies that lobbies the EC and European governments to influence policy towards the industry. An ICL spokesman says that the decisions are "unfortunate, but not devastating".

ICL, never more than a minor player in JESSI, will continue to participate in two JESSI projects in computer-aided design when the companies involved decided there was no other European company to take ICL's place. ICL researchers are also still working on a number of EC projects.

Peter Aldhous

Starting again from scratch

London

THE Iraqi occupation of Kuwait has left the scientific infrastructure of that country, formerly among the most advanced in the Arab world, in a shambles, and Kuwaiti scientists now face the unenviable task of rebuilding from almost nothing.

Restoring Kuwait's laboratories, gutted during the occupation, will take tens of millions of dollars, and the effort seems sure to be delayed until a good start has been made on the more pressing need of restoring the country's basic infrastructure. But an even more difficult task will be replacing the now-departed army of foreign scientists and technicians who made up a large proportion of Kuwait's scientific workforce. Many of those workers seem likely to stay away, at least until the political and physical environments return to normal.

Little is left at Kuwait's two important centres for research, Kuwait University and the Kuwait Institute for Scientific Research (KISR). During the occupation, Iraqi soldiers and scientists systematically looted them of equipment, books and journals, according to many reports from Kuwait (see, for example, *Nature* 349, 450; 7 February 1991). Staff from the University of Basra, Iraq, are reported to have removed thousands of personal computers from KISR. Gisli Sigurdsson, formerly of the university's teaching hospital, says that Iraqi scientists he recognized from scientific meetings were among those he saw taking equipment and teaching materials from the hospital.

Before the invasion, Kuwait's laboratories were the equal of many in North America and Western Europe, says David Britt, a former Kuwait University faculty member who has been running a newsletter from Liverpool for the university's disbanded staff. By contrast, the university laboratories Britt has visited in Baghdad were of "secondary school standard", in Western terms — which may explain why Iraqi scientists were willing to take part in the theft of Kuwait's facilities.

Most university and KISR buildings are now reported to be intact, but empty. Gary Burns, formerly of Kuwait University Medical School and now at the Free University of Berlin, says he witnessed the looting of the Medical School library. "Even the shelving" was taken, he says.

Kuwaiti scientists realize that rebuilding their laboratories is likely to take several years, as bringing back basic services, such as water and electricity supplies, and extinguishing the hundreds of fires across Kuwait's sabotaged oil fields will take precedence. Abdullatif Al-Bader, dean of the Kuwait Medical School, predicts that restoring that school to its pre-invasion state is likely to take some three years. Al-Bader, who was stranded in Montreal when Iraq invaded, expects to return to Kuwait within the next six weeks to assess the precise extent

of the damage.

The university is a more urgent concern than KISR, mainly because of its teaching function. The short-term priority, says Ali Trarh, from the Kuwait University office at the Washington embassy, is to find temporary places for as many students as possible at universities in the other Gulf states.



Devastation in Kuwait City.

Replacing journal back issues and out-of-print books will be a major problem. But the departure of the majority of university and KISR staff is a greater worry. Given the cur-

KUWAIT

Palestinians may be barred from courses

London

IN the wake of the Palestine Liberation Organization's support for Saddam Hussein during the Gulf crisis, many human rights groups are fearing a backlash against Kuwait's large Palestinian population. Palestinian students and researchers may find themselves discriminated against in the postwar rebuilding of Kuwait.

At Kuwait University, former staff members say that they fear the Kuwaitis' hostility towards Palestinians will bar many of the university's best students from completing their courses. Before the Iraqi invasion, Kuwaiti citizens were just barely a majority of the university's 17,000 students, despite the preferential treatment given to Kuwaiti applicants. Most of the rest were other Arabs, especially Palestinians.

The Palestinians tended to be the university's best students, says David Britt, formerly of the faculty of allied health, but he fears that they will not be welcomed back. With the university gutted, its students will have to find temporary places at schools in the other Gulf states, but there are too few

places to accommodate all the students, says Ali Trarh, from the Kuwait University office in Washington. In finding places for students, preference must be given to native Kuwaitis, Trarh says.

"Many students face the prospect of finding nowhere to go", says Gary Burns, formerly of the medical school. "It's desperately upsetting."

The anti-Palestinian feeling may also affect the Kuwait Institute for Scientific Research (KISR). Before the invasion, the majority of KISR staff were foreign nationals, including a large number of Palestinians. Richard Wilson, from Harvard University, a regular visitor to KISR over the past decade, fears that this memory may place KISR's restoration low on the Kuwaiti government's list of priorities. Ironically, Wilson says, the expertise of KISR's former staff would be particularly valuable in solving Kuwait's immediate problems. KISR researchers concentrated on applied research, including work into air pollution, seawater desalination and petroleum technology.

Statements by Kuwaiti officials indicate a preference for the latter course. The Kuwaiti government aims to ensure that a majority of the country's postwar population is native Kuwaiti. Before the Iraqi invasion, Kuwaitis made up only 30–40 per cent of a population of some 2 million.

Peter Aldhous

Peter Aldhous

The growth of British science

SIR — It is widely believed that British science, as a consequence of government cuts, is in decline. The strongest evidence has been adduced, bibliometrically, in the Commentaries published in *Nature*^{1–3} or in *New Scientist*⁴ by Irvine, Martin *et al.* of the Science Policy Research Unit, University of Sussex, or by Smith *et al.*⁵ of the Science and Engineering and Policy Studies Unit (SEPSU).

The bibliometricians studied the numbers of papers or citations recorded annually on the Computer Horizon Inc./National Science Foundation database, which is derived from the material collected for the *Science Citation Index* by the Institute for Scientific Information (ISI). Irvine, Martin *et al.* then summarized their findings in the titles of their papers: "The writing on the wall for British science"⁴; "Charting the decline in British science"¹; and "The continuing decline of British science"². Within, their language was strong: "a catastrophe that we face as a scientific and educational nation"¹. The SEPSU commentary⁵ was soberly titled "National performance in basic research" but its findings were grave — "UK performance has deteriorated since the early 1970s both in absolute terms and relative to other countries".

These titles and conclusions, however, are misleading. The workers used a 'fixed-year database'. They studied all the journals that ISI collated in 1973, and they continued to count the papers published in those same journals until 1982. As Britain's share of papers published in those journals fell by 10 per cent between 1973 and 1982, it appeared that Britain's science was in decline. But 1973–82 witnessed an enormous expansion of science worldwide. So many new journals were created that ISI expanded its coverage from 2,364 journals in 1973 to 3,246 in 1982⁶. *Cell*, for example, was founded in 1974, yet this was one of the new journals that the bibliometricians ignored. If this 40 per cent increase in the numbers of journals reflects a 40 per cent increase in the number of papers (not necessarily a correct assumption), then between 1973 and 1982 Britain increased its numbers of papers by some 30 per cent. Moreover, throughout the 1980s, the British came second only to the United States as the source of the 300 most-cited papers published annually (see detailed references in ref. 7). Unfortunately, neither Irvine *et al.*^{1–4} nor Smith *et al.*⁵ referred to the work of Price, who showed that the numbers of scientific journals have doubled every 15 years since 1760⁸.

Furthermore, Irvine *et al.*^{1–4} and Smith *et al.*⁵ actually found that Britain's share of publications worldwide fell from 9.5 per cent to 8.3 per cent between 1975 and 1979 (inclusive), but was then maintained at 8.3 per cent between 1979 and 1982 (inclusive). Yet Irvine *et al.* predicted renewed decline:

"the UK world share of publications may decline at a somewhat faster rate over the foreseeable future"¹ because, as Smith *et al.* stated, "the cuts in university funding initiated by the British government in 1981, for example, are unlikely to have had time to influence our bibliometric measures"⁵. Yet these predictions have not been borne out. Indeed, Irvine *et al.* have since shown that the post-1979 plateau in Britain's share has been maintained⁹. How did these bibliometricians get their predictions of further decline wrong?

In discussing financial support for science, the bibliometricians assumed that the government's cuts in science funding would shrink Britain's science. But that was not a safe assumption. Internationally, there is a good correlation between the numbers of papers published per capita by any particular country and that country's gross national product (GNP) per capita¹⁰; that is, rich countries do a lot of science, poor countries do less. But different governments employ different policies. The Swiss and Japanese governments fund only about 22 per cent of their national science; the Australian and New Zealand governments fund 64 per cent and 80 per cent, respectively¹¹ (the average for industrialized countries is 43 per cent)¹¹. yet these different policies appear to make little difference to the numbers of papers published. That is largely determined by GNP per capita — public and private money appear to displace each other.

When Irvine and Martin in their Commentary entitled "Is Britain spending enough on science?"², answered their own question with the subtitle "Britain is falling behind the major industrialized countries in its investment in academic and related research", they made two errors. First, they employed misleading titles and subtitles, because they did not actually examine Britain's expenditure on science, they examined the British government's expenditure, a very different thing; and by ignoring private support for science, they overlooked the source of growth during the 1980s. Both industrial¹², and medical charitable⁷, support for science has doubled in the past ten years, and the medical charities now outspend the government's Medical Research Council⁷. In consequence, the British universities have expanded. The numbers of full-time British university academics increased from 40,246 to 47,038 between 1976–77 and 1986–87¹³. The numbers of tenured staff fell from 32,738 to 31,432 (reflecting government cuts) but the numbers of non-tenured staff doubled from 7,508 to 15,606. There was a comparable increase in the numbers of PhD students¹³. It is these non-tenured staff (postdoctoral research workers and others on short-term contracts) who have been, in part, responsible for the continued growth of

British science. The *University Statistics* do not differentiate between those working in the humanities, natural sciences or social sciences, but the sciences apparently accounted for at least half of the 1976–77 to 1986–87 expansion.

During the 1980s, faced with static government support, the private sector doubled its support for academic science. This could have been predicted, both from international comparisons and also from British history. Britain led the world, economically, technologically and scientifically, through the Industrial and Agricultural Revolutions in the near-complete absence of state support for science¹⁴. The civic universities, the Royal Institution, Faraday, Kelvin and Davy among others, were all privately funded. The government entered on science significantly only for military reasons. The forerunner of the Science and Engineering Research Council (SERC) was created in 1916 to help develop a tank, poison gas and barbed wire. The economic justifications for SERC were invented only *post hoc*.

The displacement of public by private money carries risks. Non-governmental bodies tend to rely on short-term contracts, but they are now endowing more permanent jobs. Non-governmental bodies, it is feared, may neglect pure science, but they appear to tend to understand the importance of pure science, and they fund it¹⁵. Encouragingly, disparate private sources introduce pluralism into science funding, which may ensure a more judicious support for different projects. Moreover, private bodies compete. The Association of Medical Research Charities has recommended to its members that they now pay £6,000 a year to their PhD students. The free market has done more for PhD stipends than did years of lobbying.

In summary, British science is currently experiencing, through its growth, a gradual privatization. That carries both risks and benefits, but it is not the same as decline.

TERENCE KEALEY

University of Cambridge,
Department of Clinical Biochemistry,
Addenbrooke's Hospital, Hills Road,
Cambridge CB2 2QR, UK

- Irvine, J., Martin, B.R., Peacock, T. & Turner, R. *Nature* **316**, 587–590 (1985).
- Irvine, J. & Martin, B.R. *Nature* **323**, 591–594 (1986).
- Martin, B.R., Irvine, J., Nasin, F. & Skerritt, C. *Nature* **330**, 123–126 (1987).
- Martin, B.R., Irvine, J. & Turner, R. *New Scientist* **104**, 25–29 (1984).
- Smith, D.C., Collins, P.M., Hicks, D.M. & Wyatt, S. *Nature* **323**, 681–684 (1986).
- Evaluation of National Performance in Basic Research* ABRC Science Policy Studies No. 1 (1986).
- Kealey, T. *Scientometrics* **20**, 413–438 (1991).
- Price, D.J. de S. *Little Science, Big Science* (Columbia University Press, New York, 1963).
- Martin, B.R., Irvine, J., Narin, F., Skerritt, C. & Stevens, K.A. *Science and Public Policy* **17**, 14–26 (1990).
- Braun, T. & Schubert, A. *Scientometrics* **13**, 3–9 (1988).
- OECD Science and Technology Indicators No. 3, (OECD 1989).
- Annual Review of Government Funded Research and Development 1989* (HMSO, London 1989).
- University Statistics* (University Grants Committee, London).
- MacLeod, R.M. *Minerva* **4**, 197–230 (1971).
- Rosenberg, N. *Research Policy* **19**, 165–174 (1990).

Bayesian reasoning in science

Colin Howson and Peter Urbach

Bayesian scientific reasoning has a sound foundation in logic and provides a unified approach to the evaluation of deterministic and statistical theories, unlike its main rivals.

OURS is an uncertain world, though fortunately all things are not equally uncertain. But we are used to grading uncertainty. We are fairly certain, for example, that the textbook 'laws' of physics will remain valid for the foreseeable future, and very certain that they will remain in force through next week. We are much less certain about tomorrow's weather. We not only grade uncertainty — sometimes we measure it numerically, as in effect we do when we talk about the odds we think are merited by some predictive hypothesis or other. The more certain we are that an event will or will not take place, the longer (bigger) or shorter (smaller) the odds we are prepared to give.

Gamblers use the odds scale because odds tell you directly the proportions in which the stakes are divided after the outcome. But it is not a very good scale on which to measure uncertainty as such. Because odds are ratios the odds scale starts at 0 and is unbounded to the right (infinite odds). An equally balanced uncertainty, corresponding to 1 on the odds scale, therefore cannot be represented as a midpoint. To symmetrize, we transform odds into the scale of probabilities, using the formula probability (p) = odds/(1 + odds), and put the probability corresponding to infinite odds equal to 1. So probabilities lie between 0 and 1 inclusive, and even-money odds now become the midpoint of the probability scale, as desired.

Axioms

The formal theory of probability was born in the late seventeenth century in the work of Fermat, Pascal, Huyghens and James Bernoulli. It is summarized today in four basic laws, or axioms. The first says that the probability of any hypothesis h is a non-negative real number: $P(h) \geq 0$. The second says that the probability of any necessary truth t is 1: $P(t) = 1$ (a necessary truth is one that is true whatever the world might be like; 'if it is raining then it is raining' is an example). The third, 'the additivity principle', says that if h and h' are mutually exclusive then the sum of their probabilities equals the probability of their disjunction: in symbols, $P(h) + P(h') = P(h \text{ or } h')$. The fourth, says that the conditional probability $P(h|h')$ of h given h' , is equal to the unconditional probability $P(h \& h')$ of the conjunction h and h' , divided by the unconditional probability $P(h')$ of h' where that probability is positive: in symbols, $P(h|h') = P(h \& h') / P(h')$, where $P(h') > 0$.

Mathematical probability theory began life as theory of uncertainty. But in the late nineteenth century the probability axioms

became recognized also as the laws of objectively random phenomena, or objective chance. Chance turns out to play an essential role in modern science, in the theory of statistical sampling, information theory, demography, genetics, thermodynamics and quantum theory.

Our concern here is with the older idea of probability as the foundation of a theory of uncertainty. We shall show how some modern developments endorse this idea, and how they enable us to see the rules of the probability calculus as a logic of inductive inference. Suppose h is some scientific hypothesis. Experimental data can never conclusively prove that h is true, even if it is true. So you are never absolutely certain of h 's truth, only more or less. The inductive inference consists in assessing the degree of certainty warranted by the evidence. To the heirs of Bernoulli and Laplace, this means measuring the probability of h relative to data e . Scientists often estimate the probabilities of theories, but attempts to provide an objective basis for these estimates have uniformly failed.

But if the probability of a hypothesis merely reflects our own personal degree of belief in h , how can an objective logic of inductive inference be based on such probabilities? There is no paradox here. Your degrees of belief may be personal to you, but it does not follow that they are necessarily unprincipled or anarchic — in the first place, they must satisfy the axioms of probability. Of the many arguments that have been advanced to demonstrate this, the simplest is due to Frank Ramsey and Bruno de Finetti, who discovered it independently in the 1920s and 30s.

Their result is often called the Dutch book theorem. Consider a contract whereby one party agrees to exchange with the other a sum pS for the chance of receiving S if h is true and nothing if h is false. S is a non-zero sum of money or some other divisible good, called the stake. The payoffs to the first party thus are $S - pS$ if h is true and $-pS$ if h is false (the payoffs to the other party are of course the same with the signs reversed). The contract is tantamount to a bet in which the first party is betting on h at odds $pS:(S - pS)$, that is, odds $p:(1 - p)$. It is easy to see that p is the quantity obtained by symmetrizing the odds through the transformation $p = \text{odds}/(1 + \text{odds})$, we introduced earlier. There we called p a probability, but so as not to prejudice matters we shall now call it the betting quotient associated with the odds.

Suppose p is such that you deem the odds

$p:(1 - p)$ fair, in the sense that to the best of your knowledge there is no advantage to taking either side of the bet. It is customary to identify this value of p with your degree of belief in h . Now consider some arbitrary finite set of hypotheses h_i , where p_i are your corresponding fair betting quotients on the h_i . A betting strategy with respect to the h_i is a set of decisions of the form 'bet on (against) h_i ', for each i . Ramsey and de Finetti showed that if the p_i do not satisfy the probability axioms, then there are stakes S_i and a betting strategy for the h_i which must result in a certain loss for whoever follows that strategy (such a set of stakes is known as a Dutch book). Hence you cannot consistently maintain that the p_i are all fair if they do not satisfy the probability axioms.

Proof

The proof of the Ramsey-de Finetti result uses no more than high-school algebra. Consider axiom 2 of the probability calculus, that $P(t) = 1$ if t is a necessary truth. Suppose that $p = P(t)$ is greater than 1. Because t is necessarily true the bettor on t will make a guaranteed loss of $S - pS$. If p is less than 1 then the bettor against t will make a guaranteed loss of pS . Either way one party or the other is guaranteed to lose, and so no value for p other than 1 can be fair. It is equally simple to see why $P(h)$ must be non-negative, where h is any hypothesis, and only slightly less straightforward to see how to justify the remaining two axioms of the probability calculus.

An immediate consequence of the probability axioms is Bayes's theorem, which has given its name to this approach. Thomas Bayes (1702–1761) was an English Nonconformist clergyman, a gifted mathematician, and a fellow of the Royal Society. His seminal work on probability is contained in one short memoir published posthumously in 1763.

Bayes's theorem says that, for any propositions h and e

$$P(h|e) = \frac{P(e|h)P(h)}{P(e)} \quad (1)$$

In the usual applications of the theorem, h is some hypothesis and e the evidence against which it is to be evaluated. $P(h|e)$ is the posterior probability of h on e , $P(h)$ is the prior probability of h , and $P(e|h)$ is the likelihood of h on e . Equation 1 can be rewritten as:

$$P(h|e) \propto P(e|h)P(h)$$

That is, posterior probability is proportional

to prior probability multiplied by likelihood. $P(e|h)$ can be interpreted as the degree of determinateness with which h explains e , and $P(h)$ is your estimate of the weight of evidence in favour of h before you know e . In practice it is by balancing empirical and prior factors in just this sort of way that hypotheses are evaluated. Neither factor by itself is decisive: the empirical success may, we might feel, be merely freakish, whereas the prior plausibility needs to be checked against empirical tests.

Elements

The probability calculus does more than explain how states of belief decompose into *a priori* and empirical elements. It also shows how the factor $P(e|h) / P(e)$ in Bayes's theorem, can be analysed to explain finer structure present in our informal reasoning. To see how, we need to invoke a further simple consequence of the probability axioms; namely, that

$$P(e) = P(e|h)P(h) + P(e|\sim h)P(\sim h)$$

where $\sim h$ is the negation, or denial, of h . Substituting in equation 1, we get

$$P(h|e) = \frac{P(e|h)P(h)}{P(e|h)P(h) + P(e|\sim h)P(\sim h)} \quad (2)$$

Equation 2 tells us that $P(h|e)$ depends also on both the factors $P(e|h)$ and $P(e|\sim h)$, or how probable e would be were h true or false, respectively. This dependence is more clearly brought out in the following equivalent formulation

$$P(h|e) = \frac{P(h)}{P(h) + \frac{P(e|\sim h)P(\sim h)}{P(e|h)}} \quad (3)$$

Equation 3 says that because $P(\sim h) = 1 - P(h)$, the posterior probability of h on e depends only on $P(h)$ and on the size of the so-called likelihood ratio $P(e|\sim h) : P(e|h)$, and approaches 1 as this ratio approaches 0. It is not difficult to show that if h' is any hypothesis implying $\sim h$, that is h' is a potential alternative to h as an explanation of e , then the factor $P(e|h)$ is an increasing function of $P(e|h')P(h')$: in other words, the degree of confidence one will repose in h on the basis of e decreases with the extent to which e is explained by any plausible alternative to h . This is of course exactly the informal criterion used in practice for a strong positive evaluation of any hypothesis: no hypothesis is ever regarded as secure until there is no plausible alternative explanation of the data.

Scientific methodologies have to take a view about the kinds of conclusion scientists draw, and should provide appropriate mechanisms for those inferences. The bayesian method meets these conditions by characterizing a scientific conclusion about a hypothesis as a statement of its probability, and by providing Bayes's theorem as the mechanism for calculating that probability.

But bayesianism has been widely criticized because it is based on personal, hence subjective,

probabilities. Scientific inference, critics say, should be perfectly objective. As one leading antibayesian philosopher put it: "The cognitive value of a theory has nothing to do with its psychological influence on people's minds... [but] depends only on what objective support it has in facts".

The objectivist ideal implicit in such criticisms is immensely attractive; it would be nice if disagreements in science could be resolved by taking the contending hypotheses and impartially measuring their 'objective cognitive values'. But is there any such thing?

Karl Popper and R. A. Fisher are among those who have tried to devise a purely objective methodology. Popper's starting point was the fact that general, deterministic hypotheses (simple example: 'All swans are white') can often be decisively falsified by evidence (for example, 'this is a black swan'). And his famous thesis is that only hypotheses that are falsifiable by possible or conceivable observations are scientific. A scientific theory may also make predictions which can then be experimentally checked. If the predictions are verified, Popper calls the theory corroborated. Now this is a perfectly objective statement about the theory, but does it amount to an evaluation, or carry information about the theory's cognitive value? Popper acknowledges that the hypothesis is not conclusively proved in the corroboration process; nor can it be said that its objective probability is augmented, because, as we shall explain, no one has managed to make sense of the notion of a theory's objective probability. How then can one interpret the corroboration notion without regarding it, Bayes-like, as reflecting our enhanced subjective belief in the corroborated hypothesis? Popper sometimes says that it is rational to prefer a corroborated hypothesis over one that is not, on the grounds that it is better tested; but this turns out, disappointingly, to be just another way of saying that it is corroborated. The dismal fact is that, as Popper conceded, a theory's degree of corroboration is simply a record of its performance in the various tests, with no implication for its 'cognitive value'.

A statistical hypothesis attributes statistical probabilities, or chances, to events. These are not the subjective degree-of-belief probabilities discussed earlier, but objective properties of repeatable experiments. A simple example is the hypothesis that a particular coin is fair, that is, has equal statistical probabilities of a half of landing heads and tails. And what this is standardly taken to mean is that if you tossed the coin repeatedly, the relative frequency of heads in the resulting sequence of outcomes would tend to 0.5, as the number of throws increased to infinity.

Modern science deals extensively with statistical theories, so how they are tested and evaluated is an important question. But Popper's approach is inapplicable to statistical hypotheses, which are not falsifiable.

The question of how to test and evaluate statistical hypotheses in an objective, non-

bayesian way was taken up by Fisher in the 1920s, and his approach has developed into an influential body of doctrine, known as the classical theory of statistical inference.

Classical statistical inference has two principal parts, the first relating to the testing of hypotheses (using significance tests) and the second to estimating the values of unknown parameters. The essential principles of classical inference can be appreciated through the simplest examples. Consider again the hypothesis that this coin is fair. To test it, you might toss the coin a predetermined number of times, say 20, and count the number of heads obtained. There are 21 possibilities, ranging from no heads and 20 tails to 20 heads and no tails. For a significance test, the statistical probability of each, relative to the test, or null hypothesis, must be calculated. The classical method is then to choose a subset of the possible results — usually in one or both of the distribution's tails. Several considerations dictate this subset (or critical region); an important one is that the probability that any actual outcome of the experiment falls within it when the null hypothesis is true is fairly small; 0.05 has established itself as an acceptably small value. Finally, if the outcome obtained in the experiment is in the critical region, it is said to be 'significant at the 5 per cent level'.

Whether a result is significant at a particular level seems a perfectly objective fact. And classical statisticians believe that from it one can draw a similarly objective judgment about the null hypothesis. Hence, a result significant at the 5 per cent level is said to entitle one to 'reject the null hypothesis at the 5 per cent level'.

The meaning of this expression needs analysis, however, for although the idea of rejecting a hypothesis is fairly intuitive, the idea of doing so at some percentage level is not. Fisher sometimes wrote of the hypothesis being disproved in a significance test, though he of course knew that there can be no logical disproof of statistical hypotheses. Statisticians now agree that all one can infer for certain about a hypothesis confronted with a significant result is that either the hypothesis is true, in which case the result is relatively improbable, or it is false. Indeed, Fisher held that the "force of a test of significance" resided precisely in this dichotomy. But the dichotomy has no force, as it says nothing about the null hypothesis beyond the tautology that it is either true or false.

Advice

Another response to a significant result is due to Jerzy Neyman and Egon Pearson, who invented the currently standard form of a significance test, which is slightly more complicated than that given above, in that rivals to the null hypothesis are brought into the picture. Neyman and Pearson suggested that although we are not entitled to believe that the null hypothesis is false, we should act in our practical life as if we did believe that. This oft-repeated advice is always justified

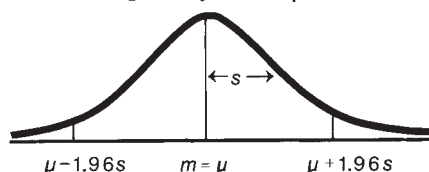
by saying that if you performed significance tests repeatedly and if each time the result was significant at the 5 per cent level you acted as if believing the null hypothesis was false, you would "in the long run" be wrong on only "around 5 per cent" of the occasions.

This argument, although superficially plausible, is mistaken in every respect. First, it is surely absurd to act as if you believed a hypothesis false, when you are uncertain that it really is. It would mean, for example, that you would willingly accept a wager in which you received £10 if the hypothesis was finally established to be false, and you had to forego all your worldly goods if it was finally established to be true. No reasonable person would accept such a bet, without being absolutely certain that the hypothesis in question really was false. Second, the justification is fallacious. It starts from the fact that the probability of rejecting a true null hypothesis equals the significance level; from this, it infers that if a test were repeatedly performed using a 0.05 significance level, the approximate frequency of rejecting a true null hypothesis would be 5 per cent, in the long run. But the inference is simply a nonsequitur: you cannot derive from the probability of an event even the approximate frequency with which that event will appear in any actual run of trials, however long. Finally, the justification is irrelevant, because it refers to the supposed frequency of drawing a wrong conclusion in a sequence of possibly imaginary tests, but says nothing about the particular case at hand.

Scientists often need to know the value of physical parameter that cannot be measured directly. The task of gauging the mean height of a very large population is a simple example. In such cases, an indirect measurement must be made and this is standardly done by examining a suitably selected sample.

Techniques

Classical statisticians have devised techniques that supposedly permit objective estimations of parameters, the principal one being that of the confidence interval, which we can explain through the example already mentioned. Let the unknown, mean population height be μ . Assume that we know the standard deviation, σ , of heights in the population. Now suppose a random sample of size n is drawn from the population. The mean height of people in such a sample is a random variable, m . Clearly m can take many possible values, some more probable than others. The distribution representing this situation is 'normal' and its standard deviation is given by $s = \sigma/\sqrt{n}$:



This distribution is a plot of possible

sample means against probability densities, not probabilities. The important fact for the present discussion is that the probability that m lies between two points is proportional to the area enclosed by them and the curve. Because the distribution is normal, it follows that with probability 0.05, $-1.96s \leq \mu - m \leq 1.96s$. Rearranging these inequalities gives the result that with probability 0.95, $m - 1.96s \leq \mu \leq m + 1.96s$. Suppose m' is the value of m that is actually observed in the experimental sample. Then, because we know σ and n , the terms $m' + 1.96s$ and $m' - 1.96s$ can be determined; the interval between them is called a 95 per cent confidence interval for μ , and classical statisticians regard such an interval as a reasonable estimate of μ .

The statement that such-and-such is a 95 per cent confidence interval for μ seems objective. But what does it say? It might be imagined that a 95 per cent confidence interval corresponds to a 0.95 probability that the unknown parameter lies in the confidence range. But in the classical approach, μ is not a random variable, and so has no probability. Nevertheless, statisticians regularly say that one can be '95 per cent confident' that the parameter lies in the confidence interval. They never say why.

In fact, there is a decisive reason why not. The confidence interval is derived from the probability distribution of sample means depicted above. This distribution gives the probabilities of all the sample means that you might have got in the experiment. It is usual to assume that all those possible samples have the same size as the actual sample. This assumption is crucial, because the shape of the distribution, and hence the width of the confidence interval, is affected by that size. Now the set of possible samples is determined by the experimenter's intention. If he had deliberately set out to sample exactly n people, the usual assumption would be justified. But suppose each time he selected a person from the population, he also tossed a fair coin, and that he planned to stop sampling as soon as the coin had produced 5 heads; or suppose the plan was simply to examine as many people as possible before lunch, or before getting bored. With any of these plans, the experimenter might still have arrived at a sample of n , but the set of possible samples would have been different, and hence, so would the confidence interval. So the degree of confidence we are invited to place in an estimate inevitably depends on the private plans of the experimenter, which is surely immensely counterintuitive.

This is the so-called stopping-rule problem. It also affects significance tests. In our earlier example, it was assumed, as it normally would be, that because the coin was tossed 20 times, all the of possible outcomes would exhibit 20 heads and/or tails. But these are the possible outcomes only if the experimenter had a premeditated plan to throw the coin 20 times. Had the plan been to stop the experiment when, say six heads ap-

peared, he could have got just the result he did, but with a different list of unrealized, possible outcomes. Because significance is calculated by reference to these possibilities, a result could be significant if the experimenter had had one plan (or stopping rule) in mind, but not significant if it was another.

This dependence of significance tests and confidence interval estimates on the subjective, possibly unconscious intentions of the experimenter is an astonishing thing to discover at the heart of supposedly objective methodologies. It is also a most inappropriate thing to find in any methodology, for the plausibility, or cognitive value, of a hypothesis, and our rational confidence in an estimate should not depend on the contents of the experimenter's mind.

Illusion

Popper's corroboration idea, and the theories of significance tests and confidence intervals were developed as supposedly objective methodologies in conscious reaction to subjective bayesianism. These methods all issue in apparently objective statements, couched in a deceptive terminology which gives the impression that an important, objective, theoretical evaluation is being achieved. But this is illusory. Corroborating a hypothesis does not strengthen it, a significant result has no significance for the truth of the null hypothesis, and a 95 per cent confidence interval has no right to impart confidence, let alone 95 per cent's worth, to an estimate.

Unlike these pseudo-objective methodologies, the bayesian approach has a solid foundation. It provides a unified approach to deterministic and statistical theories, and to questions of testing and estimation, unlike the many *ad hoc* recipes of the classical approach. It is also intuitively right. We can illustrate this by sketching the bayesian way of estimating a population mean. It starts with a distribution of subjective prior probabilities over the range of possible values of the parameter. Then, using Bayes's theorem and the sample evidence, a corresponding posterior probability is calculated. The prior probability curve typically would be very spread out, indicating considerable initial uncertainty about the parameter value, whereas the posterior probability would be concentrated in a narrow region. Then if 95 per cent of the area under the curve was enclosed between two points a and b , the bayesian estimate of the parameter would be of the form ' μ lies between a and b with probability 0.95'.

This bayesian conclusion has a clear meaning and is just the kind of conclusion people do come to. It is derived from the mean of the experimental sample alone, not the means of possible samples. Hence, it is unaffected by the experimenter's subjectively intended stopping rule, which is as it should be. Finally, the posterior distribution is very insensitive to variations in the prior distribution, and this insensitivity increases

rapidly with the sample size. Hence two people starting from very different prior beliefs must converge in their posterior beliefs as evidence accumulates. Again this seems realistic and shows that the objection to bayesianism that it makes scientific inference purely subjective is misguided.

A traditional complaint against the bayesian theory is that it introduces numerical calculations into areas where a rough and ready reliance on intuitive methodological principles seems to have worked well enough. Indeed, bayesian theory subsumes much basic methodological lore. Why go to all the trouble of writing down difficult formulas instead of continuing to do what we did well enough without them?

One answer is that 'writing down these formulas' exhibits intuitive procedures as consequences of fundamental principles of logic. Doing something correctly is one thing; to know you are doing it correctly, and why, is equally important. But there is a practical reason also. The rise of artificial intelligence and in particular the development of rule-based expert systems have made the question of what is the best method for modelling uncertain reasoning a matter of great practical urgency.

Bayesianism is only one of the mathematical theories of uncertainty currently under evaluation. The others were largely inspired by dissatisfaction with the bayesian approach. The principal alternatives are the Dempster-Shafer theory¹, incorporating a nonprobabilistic measure of belief due to Shafer, and Dempster's rule for combining evidence; 'possibility theory', put forward in the 1970s and based on fuzzy set-theory^{2,3}; and the approach, also developed in the 1970s, based on so-called certainty factors and embodied in the MYCIN and EMYCIN calculi⁴.

Uncertainty

All these theories embody numerical measures of uncertainty, and rules for revising uncertainties under the impact of new evidence. All are distinct from each other, though there turn out to be systematic relationships between probabilities, the belief and plausibility measures of the Dempster-Shafer theory, and possibility measures³. But only in the bayesian theory must the assignments of numerical uncertainty to hypotheses satisfy the probability axioms. Although in the Dempster-Shafer theory the fundamental quantities are called basic probability numbers, defined on all the subsets (roughly, hypotheses) of a frame of discernment (hypothesis space), these are not formally probabilities, nor is the belief function (Bel) constructed from them. In fact, all the other theories we have mentioned contrast with the bayesian by being nonadditive: that is, the quantities functioning as degrees of belief in each do not add over incompatible alternatives.

There is a vigorous and occasionally polemical debate under way on the relative

merits of these different approaches to modelling uncertainty. With its profusion of manifestos, it recalls an earlier controversy about the role of logic in knowledge-based computer systems. One prominent defender of the bayesian view, Peter Cheeseman, has emphasized⁵ the analogies with that earlier debate; logic, like probability, had historical precedence and was well understood, and the newer representation languages, like the newer theories of uncertainty, based their claims on alleged problems with a logic-based approach.

Controversy

Some of the controversy inevitably turns on the ease with which the various approaches can be implemented in computer systems; bayesian models are not always the most practical in these situations. They tend to have exponential informational complexity, as is also the case with Dempster-Shafer theory, depending on how many hypotheses are assigned non-zero probability. By contrast, MYCIN and EMYCIN have linear complexity. But the fundamental disagreements are philosophical, not practical, and one of the most controversial conceptual problems alleged to arise within the bayesian theory, is that of evaluating the probabilities, especially the prior probabilities, to be plugged into Bayes's theorem. In most applications there is relevant background information: in a disease diagnosis problem, for example, this would consist of the available clinical data, together with some theory. All this information should be made use of in determining probabilities. But how should this be done?

Bayes's theorem is the standard technique by which probabilities are adjusted to data (though there are generalizations of it which collectively go under the heading of probability kinematics), and the priors in any calculation can also be posterior probabilities calculated from the historical data. Ditto for the priors used in the calculation of those posteriors. But at some point in this backward progress, prior probabilities will have to be used which merely reflect opinion. This allegedly opens the doors to an unwelcome subjectivism.

To deflect the criticism, there have been several attempts to provide explicit, 'objective' rules for calculating priors. The best known is as follows: regard the hypotheses as defining subsets of some appropriate universe of 'elementary' possibilities, in the same sort of way that the hypothesis 'this die will land an even number upwards' defines the subset {2,4,6} of the set {1,2,3,4,5,6} of possible outcomes of throwing the die. Now make the prior probabilities, or probability densities if they form a continuum, of all these possibilities equal. If the universe of possibilities is finite and has m members, and if h is true in n of these, then the desired 'objective', or, as it is sometimes called, 'informationless' prior probability of h is equal to n/m .

A recent attempt to justify equal, or uniform, prior probability assignments appeals to the maximum entropy principle (MEP), according to which 'informationless' prior distributions are those which maximize uncertainty, in the sense of Shannon entropy. For in the absence of constraints other than those imposed by the probability axioms themselves, MEP prescribes uniform prior distributions.

A technical difficulty with uniform prior distributions is that they cannot properly be defined over infinite intervals, like the whole real line. But the major objection to this strategy is the arbitrariness implicit in the choice of the space of equiprobable alternatives. There are in general many different ways of embedding a set of hypotheses and data in some space or other of possibilities, and it turns out that depending on which you choose, you may get different prior probabilities for your hypotheses. And the choice is bound to be arbitrary because it is supposed to be made in advance of any empirical information (for a more detailed discussion of these matters see ref. 6).

In short, there seems to be no way of 'objectively' defining prior probabilities. But our argument has all along been that this is really no weakness: it allows expert opinion due weight, and is a candid admission of the personal element which is there in all scientific work. The inventors of 'objective' methodologies, as the statistician and philosopher I. J. Good is fond of remarking, merely sweep the personal element under the carpet. Also, there are various convergence-of-opinion theorems, which show that posterior distributions based on a lot of data are usually insensitive to priors.

The subjective bayesian theory is the only one of the various calculi of uncertainty so far to have been given an explicit and powerful theoretical justification; and the argument based on the Dutch book is only one of several. The other theories have all been proposed more or less *ad hoc*, in response to supposed difficulties in the bayesian scheme. These other approaches are not valueless; some can even be regarded as interesting generalizations of the bayesian theory. But if a clear and flexible calculus of uncertain reasoning is required, with a secure foundation in logic, then the subjective bayesian theory has no real rival. □

Colin Howson and Peter Urbach are in the Department of Philosophy, Logic and Scientific Method, The London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK.

1. Shafer, G. *A Mathematical Theory of Evidence* (Princeton University Press, Princeton, 1976).

2. Zadeh, L. *Fuzzy Sets Systems* **1**, 3–28 (1978).

3. Dubois, D. & Prade, H. *Acta psychol.* **68**, 53–78 (1989).

4. Buchanan, B. G. & Shortliffe, E. H. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristics Programming Project* (Addison-Wesley, Reading, 1984).

5. Cheeseman, P. *Proc. 9th Int. Joint conf. AI, Los Angeles* 1002–1009 (1985).

6. Howson, C. & Urbach, P. M. *Scientific Reasoning: The Bayesian Approach* (Open Court, La Salle, 1989).

The education of the very young

Nature offers itself as a clearing-house for news and information about projects aimed at the improvement of education in elementary and secondary schools, now back in fashion.

WHY is the research community worldwide just now alarmed at the quality of science education offered to young people at elementary and secondary schools? This question is not new, of course. After the launching of the first Soviet Earth satellite (called a 'sputnik') in 1957, there was an urgent demand for the improvement of science education in the United States. The National Science Foundation (NSF) quickly blessed the late Jerrold Zacharias's Physical Sciences Study Committee with the funds required for the design of an improved curriculum in physics, and quickly followed with grants for the support of projects in chemistry and biology. The objectives were to make science education at once more attractive to students and more instructive. For a time, at least, it seemed as if the objectives had been attained.

By the early 1960s, belief in the importance of deliberate curriculum development in science had spread elsewhere. In Britain, the Nuffield Foundation supported projects for the improvement of science curricula in secondary and elementary schools that remained productive for 20 years or so. There were similar developments in countries as different as Australia, France, Germany and India. The guiding principles were that science education could be made attractive to young people by letting them taste the joy of discovery familiar to people working in research, and that it could be made more instructive than it had previously been by paying careful attention to the inter-play between factual learning and conceptual understanding.

By ordinary educational standards, the effort lavished on these projects was immense. The grants supporting them may have been large, but they were small compared with the sums spent by schools on new equipment, the training of teachers and the design of examinations and other methods of assessment. That the revolution of the school curriculum engineered 30 years ago coincided with the arrival of what was often called the 'new mathematics' made it seem more natural, even respectable. By the late 1960s, in the United States and most other places, there was general contentment that science education had changed once and for all, and for the better.

So why, now, should there be a resurgence of anxiety about the character and the quality of science education in the schools? One explanation is that the rhetoric of 30 years ago was never translated into reality. It was commonplace then for the curriculum develop-

ment projects to proclaim that their aim was not once-and-for-all reform, but the founding of a continuing process. In the event, science teachers in schools wilted at the prospect of further change, sponsors of the reforms discovered other ways of spending money and, in many places, schools were overwhelmed by social and structural problems — drugs, for example, and (in Britain) the overdue loosening of a rigid examinations system.

But there is more than that to say. In the United States particularly, it seems at last to have been recognized that international competitiveness hangs on the quality of school education. (Japan and Germany, the economic successes of the past half-century, run school systems which are admired, and which are also almost rigidly selective.) In Britain, the new anxiety about science education in the schools is tinged with concern about the public understanding (or lack thereof) of public issues with a technical component, from the safety of nuclear power stations to the role of *in vitro* fertilization in the treatment of infertility. And there are emerging serious shortages of people skilled in important technical fields.

So why not turn the clock back 30 years, and mount another revolution? There are several reasons why this simple stratagem would not suffice. For one thing, there are no funds (even though, in the United States, the education budget of the NSF is likely to be increased substantially this year). But now, anxiety about the quality of school education is general; science education is only a part of it. And experience has shown that the upheaval of the 1960s touched only a small proportion of the school population and that it did not contain the means for its own regeneration. Something different is needed now.

That seems to be the reason why, in the United States especially, there has been an upsurge in the past few years of less formal attempts by academics and researchers to mount collaborations with local school systems aimed at improving the skills of science teachers and stimulating an interest in science among their students. One biologist friend, for example, seems reconciled to spending the bulk of his sabbatical year in working on the improvement of the science curriculum (and the associated equipment) in his university city's schools.

Better known is the involvement of Fermilab (the high-energy physics institute west of Chicago) in school science education in Illi-

nois. Largely because of the enthusiasm of Dr Leon Lederman, director of Fermilab from 1979 to 1990, there are summer internships for science teachers and weekend programmes for their students, while the laboratory has close links with science teaching at the nearby special high school. Now, other national laboratories are following suit, as are many university departments, while the Bahcall report on astronomy in the United States, published two weeks ago, makes a special plea that the importance of astronomy as a means of educating the young should be acknowledged.

In Britain, activities such as these are less conspicuous. But several scientific societies have recently taken a selfconscious interest in school education (and could do more); the Royal Institution (in London) continues to provide 'master classes' in mathematics and a range of general services for teachers and their students at publicly maintained schools.

Much effort is being spent on the definition of what students 'should' know and understand at what age, but there is also a sprinkling of attempts to develop special curricula. One is the Understanding Science Project run by D. E. P. Hughes, based at Westminster School in London, with support from the Leverhulme Trust. This is aimed at giving students in the two senior years of British secondary schools not otherwise studying science subjects a general understanding of the kinds of technical issues that will (or should) perplex them in adult life.

The project consists of a computer-based statistics course, an instructional unit on measurement and a series of teaching modules centred on potentially contentious issues — the safety of nuclear power and the AIDS epidemic. The teaching materials will be tried out in a number of schools and colleges (more are needed) starting in September. The Schools Examinations and Assessment Council has agreed that the course will be part of the new pattern of school-leaving examinations.

There are good reasons why schemes such as this deserve to be better known. Often it will stimulate others to similar ends merely to know that they exist, but there may also be opportunities for collaboration. For this reason, *Nature* will be glad to learn of such schemes, and will be prepared to publish brief descriptions of them together with the name and address of a person to whom enquirers may write for further information.

John Maddox

Young Earth like Venus?

Stuart Ross Taylor

SURROUNDING the iron-rich core of the modern Earth is a thick rocky mantle which is overlain by a thin layer of chemically distinct crust. Yet the Earth formed through the accumulation of smaller rocky bodies, called planetesimals, which formed out of the nebula that gave birth to the whole Solar System. What happened in between? Galer and Goldstein propose¹ an iconoclastic picture in which the mantle of the young Earth developed a basaltic crust (which has since disappeared) similar to that now found on Venus (Fig. 1), but chemically unlike the basalts found on the ocean floor and quite distinct from modern continental crust. One consequence of this picture is that present-day plate-tectonic processes could not have occurred on the early Earth.

The oldest preserved rock on Earth is the Acasta gneiss, 3,960 million years (Myr) old², in the Northwest Territories of Canada (Fig. 2). Rocks formed before that are generally thought to have been destroyed by meteorite bombardment, which persisted until about 3,800 Myr ago. In preparing a chronology for the early Earth, the next early secure date comes from the Moon. There is hard geochemical evidence that the Moon melted during formation, and the age of $4,440 \pm 20$ Myr records the crystallization of the lunar crust. The earliest datable episode in the solar system is the formation of the refractory inclusions, preserved for example, in the Allende meteorite, which thus provides a convenient marker for the 'beginning' of the Solar System, or T_0 , at 4,560 Myr ago. Thus the existence of the lunar crust places an upper limit of about 100 million years after T_0 for the formation of the Earth–Moon system.

The Earth grew in this interval through the accumulation of a hierarchy of rocky planetesimals in a gas-free environment (violent solar winds had quickly flushed volatile gases from the original nebula). The Moon was probably formed at this time by a collision between the Earth and a body a little larger than Mars: the Moon is probably made from the rocky mantle of the impactor.

This history of accretionary collisions with planetesimals would have heated the early Earth so that "deep-seated global melting seems unavoidable"³. Certainly the lunar-forming event on its own was sufficiently energetic to have accomplished this. What happened next is the subject of vigorous debate⁴. Several things did not happen. There was no terrestrial anorthositic crust, analogous to the lunar crust, in part because the Earth is less well endowed with Ca and Al than the Moon, so that plagioclase — the main, Ca-rich component of anorthosite — could

not have been an early crystallizing phase from a terrestrial magma ocean: plagioclase floated on the bone-dry lunar magma ocean, but would have sunk and reacted in the wet terrestrial melt.

Also, there is no good evidence of an early granitic crust. Among the enduring myths in the geological sciences, that of a primordial world-encircling crust of 'sial' or granite shows a surprising persistence. But, such

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

FIG. 1 The crumpled basaltic surface of Venus (part of the Freyja Montes region), perhaps an analogue for the crust of the early Earth. (Magellan image P-37138).

models were driven by false analogies between basaltic magmas, which crystallize to produce a silicic residuum, and the early molten Earth. The lack of evidence for such a crust seems overwhelming⁵, and granitic crust, probably a rock type unique to the Earth, is the end product of at least three stages of derivation from a primitive mantle composition. The process is certainly inefficient, for the Earth has transformed less than 0.5 per cent of its volume into granitic continental crust in over 4,000 million years. The Moon, in contrast formed its highland feldspathic crust, about 12 per cent of lunar volume, in a few million years.

By whatever mechanism the mantle crystallized, evidence is accumulating, Galer and Goldstein argue¹, for some extensive early differentiation. The evidence comes from the geochemistry of Sm and Nd isotopes. This view is based on the assumption that the Earth's total abundances in rare-earth elements have the ratios observed in type 1 carbonaceous chondritic (or CI) meteorites, which are taken as representative of the solar nebula⁶. The Earth and the other inner planets do not, of course, have pristine CI compositions, being comparatively depleted in volatile elements relative to refractory ones. Also, the Earth's bulk Mg/Si ratio may be 25 per cent larger than the chondritic value, if the well-established upper mantle value is representative of the whole mantle (not a wholly accepted premise). So it is

possible that the Earth's rare-earth-element pattern is also non-chondritic.

Nevertheless, Galer and Goldstein assume that the terrestrial rare-earth pattern does resemble that of CI, that the Earth has chondritic initial Sm/Nd ratios, and that observed deviations are due to subsequent internal fractionation. On this basis, sufficient data have now accumulated for the authors to demonstrate that the silicate mantle in the Archaean Eon (before 2,500 Myr ago) was depleted in light rare earths (LREE). How early is less certain, but it must have been well before the time of formation of the oldest surface rocks, and possibly soon after the accretion of the Earth itself. Many terrestrial Archaean rocks are derived from such a LREE-depleted (Sm/Nd enriched) reservoir, illustrating that the depletion was a planetary-wide affair. This is demonstrated not only for a wide variety of rocks derived by mantle melting, but also by measurements on apparently preserved remnants of the Archaean mantle⁷.

The straightforward explanation is that a large portion of the upper mantle, at least, underwent partial melting, involving the extraction of the LREE and leading to an increase in the Sm/Nd ratio in the mantle. How large? The amount of Nd involved is about as much as is contained in the present continental crust (around 30 per cent, at least, of the mantle inventory) so that the event or events were on a global scale, extending to depths of hundreds of kilometres.

What mechanisms exist to deplete the upper mantle in the LREE? Extraction of an early crust is the obvious possibility. But if some sort of early crust was produced, what was its composition and what has happened to it? As noted above, early granitic or anorthositic crusts are unlikely. The best candidate is a basaltic crust (modern ocean crust is basaltic), as suggested by several other workers. Among the few features common to Venus, Earth and Mars is their habit of producing massive amounts of basaltic lavas through partial melting of their silicate mantles. As the primary melt from silicate mantle, only basalt is produced in sufficiently voluminous amounts to be a viable candidate to cause a massive depletion of elements in the mantle.

Solid silicate mantles melt because of the accumulated heat produced by radioactive elements. As these elements are themselves partitioned into the melt, they are transferred to the surface, reducing the heat generated internally, and eventually shutting the heat engine down, as occurred on the Moon over 2,000 Myr ago. Thus alkali basalts, rich in K, U and Th and the other 'incompatible' elements, can be expected to be erupted, as occurs now on Venus⁸. Galer and Goldstein postulate that the early terrestrial crust was indeed composed of such material. The LREE would also have been preferentially

enriched in the melt phase, and removed from the mantle, leaving behind a residue depleted in the LREE with enhanced Sm/Nd ratios. These will lead to an increase in radiogenic ^{143}Nd from the decay of ^{147}Sm in the mantle. This is what Galer and Goldstein consider to have happened at an early stage on the Earth.

This basaltic crust has now vanished, presumably either by subduction as happens to the present (non-alkali) oceanic crust, or by sinking into the mantle, aided by a transformation into dense garnet-bearing eclogite. This raises the question of whether plate tectonics in the modern sense, operated at this very early stage of Earth history. Galer and Goldstein argue against the notion, preferring a picture of static crust resembling the venusian case (although there is much evidence on Venus for lateral tectonics, pushing and crumpling the crust^{8,9}). The principal reason for their objection to plate tectonics is that removal of LREE from the mantle requires only a small degree of partial melting (the authors estimate 2–10 per cent). Modern plate tectonic regimes typically cause partial melting on a larger scale (over 20 per cent) and the high temperatures of the Archaean Eon would have allowed even more melting if tectonic processes occurred. Present-day plate tectonics, it seems, with the development of linear tectonics and island arcs, started only after the development of massive areas of continental crust in the late Archaean^{10,11}.

A nagging problem with all models involving an early alkali basaltic crust enriched in incompatible elements is what happened to it? It seems to have vanished, like the Chesire Cat, leaving only an enigmatic signature of early mantle depletion. Galer and Goldstein surmise that it was gradually returned to the mantle, without imparting its enriched LREE signature to rocks subsequently

derived from that region. What seems certain is that the evolved siliceous rocks (the tonalite–trondjemite suite), which built up most of the scattered regions of continental crust in the Archaean, were not derived by remelting of Galer and Goldstein's LREE-enriched basaltic crust. Although tonalites and trondjemites are generally supposed to be derived through the remelting of basalt (transformed to eclogite) in the mantle, their source was depleted, not enriched.

If the Earth was covered with an alkali basaltic crust over 4,000 million years ago, what did the surface look like in the first 500 Myr (or in the 'Hadean' Eon, a term due to Preston Cloud)? The best analogy would seem to be the present surface of Venus, now revealed in stunning clarity by the Magellan radar (Fig. 1). The differences lie in the absence of surface water on Venus, so that Venus has never been able to produce a more extensive differentiated crust of tonalitic or granitic rocks.

Of course, this whole suggestion depends on a chondritic-type bulk Earth. Perhaps we

should entertain more seriously the matter of a non-chondritic rare-earth pattern for the Earth. This of course has wider implications, constraining, for example, the width of the zone from which the Earth-forming planetesimals came. This raises entirely new and other interesting questions about the entire early inner Solar System. □

Stuart Ross Taylor is in the Department of Nuclear Physics, Australian National University, PO Box 4, Canberra ACT 2601, Australia.

1. Galer, S. J. G. & Goldstein, S. L. *Geochim. cosmochim. Acta* **55**, 227–239 (1991).
2. Bowring, S. A., Williams, I. S. & Compston, W. *Geology* **17**, 971–975 (1989).
3. Stevenson, D. J. *Nature* **335**, 588 (1988).
4. Newsom, H. E. & Jones, J. H. (eds) *Origin of the Earth* (Oxford University Press, 1990).
5. Taylor, S. R. *Nature* **346**, 608–609 (1990).
6. Anders, E. & Grevesse, N. *Geochim. cosmochim. Acta* **53**, 197–214 (1989).
7. Collerson, K. D. et al. *Nature* **349**, 209–214 (1991).
8. Basilevsky, A. T. & Head, J. W. A. *Rev. Earth planet. Sci.* **295**–317 (1988).
9. Campbell, D. B. et al. *Science* **251**, 180–183 (1991).
10. Taylor, S. R. & McLennan, S. M. *The Continental Crust: Its Composition and Evolution* (Blackwell, Oxford, 1985).
11. Krogstad, E. J. et al. *Science* **243**, 1337–1340 (1989).

CANCER GENES

Telling changes of base

Adrian L. Harris

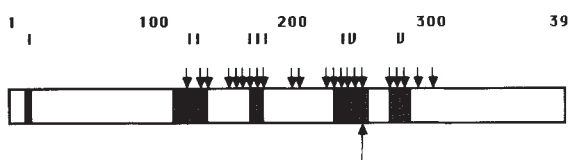
MUTATIONS in the nuclear phosphoprotein p53 are emerging as the commonest genetic change in human cancer. But perhaps more surprisingly, the p53 mutations for one type of cancer seem to be clustered in the same codon. This is the finding reported in this issue by Hsu *et al.*¹ on page 427 and by

fere with the normal p53 protein and also form a complex with heat-shock cognate protein 70 (hsc70).

Normal p53 protein is believed to be a tumour-suppressor gene involved in controlling the entry of the cell into the S phase. Mutations in p53 have been found in more than 30 different codons, whereas, for example, mutations at one of only a few specific sites are sufficient to activate *ras* to become an oncogene (such as codons 12,13,61). But of the 16 hepatomas analysed by Hsu *et al.*¹, transversion in codon 249 of G to C or to T occurred in all eight of the cases with p53 mutations, and of the five hepatomas with

p53 mutations analysed by Ozturk *et al.*², four had G-to-T substitutions, with three of these occurring in codon 249. These base changes are produced by a range of carcinogens including aflatoxin, which is partly responsible for liver cancer in China and South Africa. For this reason, it would be of considerable interest to compare cases originating in the United States and Europe.

Mutations in p53 may be markers for specific carcinogens in a variety of cancers. In a study of lung cancer, 15 out of 25 mutations were found to involve a G-to-T transversion^{3,8}, a pattern that could be produced by benzo[a]pyrene, a carcinogen in cigarette smoke. But in these cases the mutations occurred in 16 different codons (including 249) and only one codon was represented four times. The new hepatoma work is the



Schematic p53 molecule showing conserved domains I–V. The sites of mutations in sporadic tumours other than hepatoma are shown above the molecule and the hot-spot for hepatomas below.

Ozturk *et al.*² on page 429. The groups looked at hepatocellular carcinomas in patients from China (Hsu *et al.*) and southern Africa (Ozturk *et al.*), and found that in nearly every patient with a mutant p53 gene, the mutation lay in the same codon. The result contrasts with all previous studies of human cancers.

Mutations in p53 have been assessed by DNA sequencing and immunochemistry³. Over half of breast⁴, lung⁵, colon⁶ and bladder tumours (D. Neal, personal communication) have been found to carry a p53 mutation. Mutations are clustered in four regions that are highly conserved among vertebrates⁷, involving exons 5–10 (see figure). The mutations alter the half-life of the protein from about 15 minutes to several hours, which enables it to form a complex and inter-



FIG. 2 A zircon from the oldest rocks in the world, the Acasta gneisses from the Slave province of northwest Canada, 3,960 million years old. (From ref. 2.)

first time such a high frequency hot-spot has been reported.

The presence of a hot-spot in one particular type of cancer implies that there is a selective advantage for the specific mutation. Because the hepatitis B virus is also implicated in the development of hepatomas in southern Africa and China, it could be that the interaction of a virally encoded protein with the specific mutant p53 provides a growth advantage in hepatomas. It is widely recognized that transforming viral proteins such as large T antigen and E1B can interact with non-mutated p53 and that this association is part of the mechanism by which the cells are transformed. Although it is generally assumed that mutations in p53 result in a loss of function, the reports by Hsu *et al.* and Ozturk *et al.* of a mutational hot-spot at such high frequency imply an additional gain in function. Indeed, different mutations in p53 do induce different functional effects.

Most p53 mutants lose their ability to suppress the transforming effect of *ras* in co-transfection experiments, are strongly transforming themselves and form a complex with hsc70. A mutant p53 has been identified that has only the first property⁹. This implies that there are probably gain-of-function mutants. Additionally, transfection of a mutant p53 gene into cell lines lacking both p53 alleles can confer a growth advantage on the line¹⁰. The high frequency with which the mutant p53 protein is expressed and the fact that the wild-type remaining allele is lost by chromosome deletion⁶ also implies that there is a selective advantage in retaining the mutant protein. If transformation can arise by the formation of a complex between wild-type and mutant p53, then once the chromosome carrying the wild-type allele has been lost, there should be no further advantage in retaining the mutant protein.

A similar finding of tumour-specific mutations has been reported for malignant gliomas, where the same gene deletion occurs in the region corresponding to the external ligand-binding domain of the receptor for epidermal growth factor¹¹. There may well be other tumour types where the intracellular signalling system or microenvironment exerts a strong selective pressure for a particular pattern of mutation which is not seen in the same oncogene expressed at other sites.

The newly described mutational hot-spot

in p53 occurs in the middle of the region involved in the Li-Fraumeni familial cancer syndrome (codons 245–258). These mutations seem to have a different effect in the syndrome than they do in most common cancers, in that they do not lead to overexpression of mutant protein. The Li-Fraumeni lesions could therefore be loss-of-function mutations alone.

The hepatoma p53 mutants provide strong candidates for further investigation both of gain of function and of the proteins with which they interact. Expression of recombinant proteins and their use in affinity purification columns will enable the com-

parison of extracts from hepatomas induced by hepatitis B with other hepatomas and other types of tumour. The result of transfecting these specific mutated genes into different cell lines, in particular hepatoma cell lines, will be awaited with interest. This mutant protein may also provide a further link in understanding the mechanisms of viral carcinogenesis if interactions with hepatitis B viral protein products can be demonstrated. □

Adrian L. Harris is at the Molecular Oncology Laboratory, Imperial Cancer Research Fund, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK.

CELL BIOLOGY

Microtubules get the chop

Peter J. Hollenbeck and William L. Dentler

In a report in *Cell*¹, R. D. Vale describes the identification of a cell cycle-regulated factor responsible for severing microtubules. This finding may lead to a deeper understanding of the control of the cytoskeleton, the system of dynamic protein filaments which in turn determines the movements, shape and mechanical properties of eukaryotic cells.

Two of the polymers that comprise the cytoskeleton — actin filaments and microtubules — each have numerous associated proteins which are thought to modulate their stability by affecting the equilibrium between monomers and polymer. In particular, a dizzying array of proteins can associate with actin filaments; some of them stabilize the polymer through lateral interactions or capping of the filament ends, while others promote polymer disassembly by severing the filaments along their length, sequestering monomers in assembly-incompetent form, or both². Microtubules, too, have stabilizing proteins that bind laterally along the polymers³ and proteins that cap their ends⁴. Polymer severing has not been regarded as a significant component of microtubule regulation in most cells, despite evidence that it takes place in ciliary and flagellar microtubules⁵. However, Vale now demonstrates that microtubules are severed extensively *in vitro* by an activity present in extracts of meiotic amphibian oocytes, but absent from interphase extracts, thereby showing that this aspect of microtubule regulation may itself be regulated by the cell cycle.

Light microscopy

Vale developed a quantitative assay in which fluorescently labelled, drug-stabilized microtubules were observed *in vitro* by light microscopy before and after exposure to stage-specific cell extracts. Extracts from the cytoplasm of electrically activated, 'interphase' frog oocytes had no effect on the microtubules, but extracts from oocytes blocked in the second metaphase of meiosis caused a rapid and dramatic change in the microtubule population — the number den-

sity increased fivefold while their average length decreased twentyfold. Direct observation of the process in real time with video-enhanced interference optics showed that microtubules were indeed cut into shorter segments. This was unlikely to be the result of non-specific proteolysis of tubulin, the subunit protein of microtubules, because endogenous oocyte tubulin actually assembled into microtubules while the exogenous microtubules, assembled from brain tubulin, were being fragmented.

Calcium treatment

To address the question of how the severing activity might be regulated, Vale treated meiotic extracts with calcium to reduce the amount of cyclin (a protein that regulates entry into mitosis⁶) to interphase levels. The resulting cyclin-depleted extract had lost its original severing activity. In the converse experiment, addition of cyclin to interphase extracts caused them to acquire severing activity in the absence of any protein synthesis. By analogy with other proteins regulated during the cell cycle, Vale suggests that the severing activity may be controlled by protein phosphorylation, perhaps by the *cdc2* kinase activity regulated by cyclin. Unpolymerized tubulin inhibited severing activity, which may reflect the operation of a combined severing and monomer-sequestering activity similar to that of some actin-severing proteins². In addition, the presence of microtubule-associated proteins or the bundling of microtubules inhibited the severing activity, an observation that could explain the relative stability of bundled microtubules in the mitotic midbody relative to those in the chromosome-to-pole region of the mitotic spindle⁷.

The severing factor seems unlikely to be a capping protein, because tubulin from the severed microtubules reassembled with oocyte microtubules; the stoichiometry of the factor to tubulin would provide important evidence on this point. Flagellar severing does not involve microtubule capping, since

1. Hsu, I.C. *et al.* *Nature* **350**, 427–428 (1991).
2. Bressan, B., Kew, M., Wands, J. & Ozturk, M. *Nature* **350**, 429–431 (1991).
3. Gannon, J.V., Greaves, R., Iggo, R. & Lane, D.P. *EMBO J.* **9**, 1595–1602 (1990).
4. Bartek, J. *et al.* *Int. J. Cancer* **46**, 839–844 (1990).
5. Iggo, R., Gatter, K., Bartek, J., Lane, D. & Harris, A. *Lancet* **335**, 675–679 (1990).
6. Baker, S. J. *et al.* *Cancer Res.* **50**, 7717–7722 (1990).
7. Soussi, T., de Fromental, C. C. & May, P. *Oncogene* **5**, 945–952 (1990).
8. Chiba, I. *et al.* *Oncogene* **5**, 1603–1610 (1990).
9. Halevy, O., Michaelovitz, D. & Oren, M. *Science* **250**, 113–116 (1990).
10. Chen, P.-L., Chen, Y., Bookstein, R. & Lee, W.-H. *Science* **250**, 1576–1580 (1990).
11. Humphrey, P. A. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **87**, 4207–4211 (1990).

both basal bodies and isolated flagella can assemble new microtubules from their exposed ends. Little is known about specific proteins that cap microtubules, although flagellar microtubules have plus-end capping structures⁴, and one protein associated with the capping structure in *Tetrahymena* is also found on mammalian kinetochores, which bind microtubule plus ends⁸. Minus-end capping proteins also remain to be identified, but the minus ends of microtubules of the basal body are blocked from nucleating microtubule assembly *in vivo*, and factors that prevent minus-end microtubule growth are present in extracts of *Xenopus*⁹ and clam¹⁰ oocytes.

Although Vale's report is the first identification of a cell cycle-regulated microtubule severing factor, microtubule severing does occur during the excision of flagella: each of the nine doublet microtubules are cut between the basal body and flagellum proper⁵. The mechanism is unknown, but severing in the alga *Tetraselmis* is accompanied by the phosphorylation and contraction of centrin, a calcium-binding protein¹¹ found in the transition region and rootlets¹². Other factors probably act in concert with centrin to mediate flagellar excision, because a mutant form of *Chlamydomonas* is unable to shed its flagella despite containing centrin¹².

Sanders and Salisbury were the first to suggest that a process similar to flagellar severing might act to reorganize the cytoplasmic microtubule array at the transition from interphase to mitosis¹². They pointed out that proteins related to centrin are commonly present in the centrosomal region of eukaryotic cells¹³, and proposed that severing might separate microtubules from their organizing centres, an idea echoed by Vale. The availability of a quantitative assay for severing should now allow the purification of the protein or proteins responsible, and we will probably know soon whether the activity reported by Vale involves centrin or represents a distinct mechanism for severing microtubules. □

Peter J. Hollenbeck is in the Department of Anatomy and Cellular Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. William L. Dentler is in the Department of Physiology and Cell Biology, University of Kansas, Lawrence, Kansas 66045, USA.

- Vale, R. D. *Cell* **64**, 827-839 (1991).
- Pollard, T. D. & Cooper, J. A. *Rev. Biochem.* **55**, 987-1035 (1986).
- Olmstead, J. B. *Rev. Cell Biol.* **2**, 421-457 (1986).
- Dentler, W. L. in *Ciliary and Flagellar Membranes* (ed. Bloodgood, R. A.) 31-64 (Plenum, New York, 1990).
- Blum, J. J. *J. theor. Biol.* **33**, 257-263 (1971).
- Nurse, P. *Nature* **344**, 503-508 (1990).
- Gorbsky, G. J., Simerly, C., Schatten, G. & Borisy, G. G. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6049-6053 (1990).
- Miller, J., et al. *J. Cell Biol.* **110**, 703-714 (1990).
- Gard, D. L. & Kirschner, M. W. *J. Cell Biol.* **105**, 2191-2201 (1987).
- Suprenant, K. A. *Cell Motil. Cytoskel.* (in the press).
- Martindale, V. E. & Salisbury, J. L. *J. Cell Sci.* **96**, 395-402 (1990).
- Sanders, M. A. & Salisbury, J. L. *J. Cell Biol.* **108**, 1751-1760 (1989).
- Barron, A. T. & Salisbury, J. L. *J. Cell Biol.* **107**, 2669-2678 (1988).

OBITUARY

Edwin Herbert Land (1909-1991)

In 1957, Harvard University awarded an honorary doctorate to a man who dropped out of its freshman physics class of 1926. After leaving Harvard, he had continued his education in the New York Public Library and had become one of the great American scientific entrepreneurs, to be counted alongside Edison and Bell. Such was Edwin Land, who died on 1 March.

By 1932, Land had succeeded in align-

Polaroid Corporation in 1982.

Much less known than his cameras, but probably of more significance, was Land's secret work on high-altitude optical surveillance systems. In the 1950s, he led the team that designed the high-resolution camera for the U2 spy plane — the forerunner of later systems mounted in satellites.

For the last 35 years of his life, Land's main scientific obsession was colour vision, a subject that has been a seductive and passion-provoking mistress to so many physicists. His writings and lectures attracted (thoroughly justified) hostility from established colour scientists, but he was able to leave generations of undergraduates with the notion that he had refuted existing theory.

It was in the late 1950s that he captured public imagination with his two-colour demonstrations. A black-and-white photograph of a scene is taken through, say, a red filter, and a second is taken of the same scene through a green filter. A positive transparency is prepared from each of the two negatives. The first transparency is then projected on a screen through a red filter; the second is projected in register, but with

Double take — Land demonstrating his instant photographs in 1947.

ing submicroscopic crystals of iodoquinine sulphate and embedding them in a sheet of plastic, so as to yield large sheets of artificial polarizer. With his Harvard physics instructor, George Wheelwright, he set up the Land-Wheelwright Laboratories in 1932; and he established the Polaroid Corporation in 1937. Polaroid filters, for visible and infrared light, were soon being used in cameras and sunglasses, and in wartime rangefinders and night-adaptation goggles.

Land's first instant camera, yielding a picture in 60 seconds, was demonstrated to the Optical Society of America in 1947. By Christmas of 1948, it was on sale to the public. An instant colour camera was marketed in 1963. It was a classic case of giving away the razor and making the money on the blades. Land and his company became very rich. But an instant movie system, placed on the market just as domestic video systems were becoming available, proved commercially disastrous; and Land retired from the

unfiltered white light. We might expect to see an array of reds, pinks and whites — reds where the first transparency was light and the second dark, pinks where the transparencies had similar density, and whites where the second was the lighter. In fact, the image on the screen exhibits a much richer gamut of colours, which includes blues and greens.

The two-colour demonstration shows that there is no fixed relation between the spectral composition of light and the hue that we perceive at a local point in the scene. Rather, we judge colours by the company they keep. Land took his finding to contradict what he held to be the classical colour theory of Young and of Helmholtz.

The less insightful of his critics complained that his effects, apart from being rediscoveries, were all a matter of colour contrast. Those who thought more deeply, such as M. Woolfson (*IBM Journal* **3**, 313, 1959) and D. Judd (*J. opt. Soc. Am.* **50**, 254; 1960) grasped that the Land effects, like several other contrast ►

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

phenomena, were by-products of the visual mechanism that gives us colour constancy — our ability to recognize the permanent colours of objects despite large changes in the spectral composition of the illumination. In the example given above, the illumination will be regarded as pink, and patches that reflect less long-wave light (those illuminated mainly by the second transparency) will be interpreted as greenish or bluish objects.

Rather more slowly than Woolfson and Judd, Land himself realized the relationship between his two-colour demonstrations and colour constancy, and during the next two decades he developed a second series of demonstrations, the 'Mondrian' demonstrations. These were grand (but increasingly laboured) demonstrations of colour constancy. Land always eschewed the term 'colour constancy' and declined to relate his demonstrations to a long-established tradition in colour science. Yet Young himself wrote in 1807 "...when a room is illuminated either by the yellow light of a candle, or by the red light of a fire, a sheet of writing paper still appears to retain its whiteness". And in fact Land's death coincides with the bicentennial of the masterly paper by Monge, in which that great mathematician relates coloured shadows to colour constancy and draws Land's conclusion — that we respond not to the absolute properties of the stimuli that strike our retinae but to their ratios.

It must nevertheless be said that Land's attentions revived a quiescent field. And his theory of colour constancy was important. The idea was that each of the three cone systems extracts the spatial pattern of light and dark as seen by those cones, scaling each local signal according to the total range of illuminations that it finds over a larger region. Later, a comparison is made of the three separate lightness signals for a given local area, and this comparison gives the colour of the corresponding surface. This 'retinex' theory was an early example of a computational model of how the brain might accomplish a particular task.

In public manner, Land was at once shy and imperious. Like many in his position, he gained the reputation of a recluse. In 1980 he saw the potential of a vacant lot beside the Charles River, and built there his Rowland Institute. It was a cross between an art gallery and the private laboratory of a nineteenth-century gentleman scientist — with perhaps just a touch of Hearst Castle.

J. D. Mollon

J. D. Mollon is in the Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, UK.

Methane's sinks and sources

Paul J. Crutzen

ATMOSPHERIC concentrations of methane, one of the most important trace gases (both a greenhouse gas and a player in the chemistry of ozone destruction), are currently increasing by about 1 per cent a year. The concentration arises from a balance of the rate of input, increasing owing to human activity, and the rate at which it is removed. On page 406 of this issue¹, Vaghjani and Ravishankara show that the rate at which methane reacts with hydroxyl radicals has been overestimated by up to 25 per cent. As this is one of the main sinks for atmospheric methane,

constituted about 0.8 parts per million by volume (p.p.m.v.) of the atmosphere⁴. The rise to its present average atmospheric mixing ratio, about 1.7 p.p.m.v., started about 300 years ago. As methane concentrations had previously been virtually constant, the inescapable conclusion is that the increase has been caused by human activities.

It is easy to identify several human activities that have contributed to the growth in atmospheric methane. The gas is produced biologically through the anaerobic decay of organic matter. As well as occurring in natural wetlands, this happens in rice fields, landfills and the rumen of cattle. Escapes of methane from coal mines, and leaks in natural gas distribution systems and oil and gas wells are other anthropogenic sources. Furthermore, methane is produced during the burning of biomass, mostly in the tropics and subtropics. Quantifying the individual contributions, however, is not easy.

The main sink for atmospheric methane is its reaction with OH radicals. As the average global OH concentration can be estimated to within 20 per cent from the known loss of industrially produced methylchloroform (CH_3CCl_3), the total sink of methane can likewise be well quantified⁵. But this depends on the reaction rate between OH and methane. Until recently the rate constant was thought to be well established. But the painstaking analysis by Vaghjani and Ravishankara in this issue¹ shows otherwise: the authors show the accepted reaction rate to be up to 25 per cent too large. Consequently, the total quantity of methane that reacts with OH must be about 420 Tg yr⁻¹, rather than about 520 Tg yr⁻¹, as previously estimated.

With this information, one can reassess the various sources and sinks of methane, an important task if greenhouse warming is to be tackled. The loss of methane by reactions with OH is estimated to be 420±80 Tg yr⁻¹. Additional atmospheric losses involve the uptake of 30±15 Tg yr⁻¹ by soils and stratospheric reactions with Cl and O(¹D) with 10±5 Tg yr⁻¹. The total atmospheric sink of methane is, therefore, 460±100 Tg yr⁻¹. With an annual increase in the atmospheric loading by 45±5 Tg, the total methane source is thus 505±105 Tg yr⁻¹. Of this about 18–24 per cent does not contain radioactive ¹⁴C, so that about 100±20 Tg yr⁻¹ come from fossil fuels and methane hydrates. Of this, about 25±5 Tg yr⁻¹ seem to escape from coal mines. Methane hydrates are estimated to release only 5 Tg yr⁻¹, although an upper limit 100 Tg yr⁻¹ has been quoted by Cicerone and Oremland⁶. (Such a large release rate seems unlikely, however, in view of the 5°C global warming at the end of the last Ice Age, which far exceeds the warming over the past 300 years, 0.6°C: if the release rate is now 100 Tg yr⁻¹, the warming then

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

BP gas terminal at Easington, Yorkshire, tapping Britain into the natural reserves of the North Sea. If leakages are excessive, natural gas could exacerbate, not alleviate, global greenhouse warming.

the result requires a careful reconsideration of the methane budget.

Methane's influence is wide ranging in atmospheric chemistry. In the background troposphere, methane is involved in photochemical reactions that determine the concentrations of ozone (O_3) and hydroxyl (OH), the 'detergent' of the atmosphere, responsible for the removal of almost all gases that are produced by natural processes and human activities². Methane also influences the chemistry of the stratosphere: its oxidation is an important source of stratospheric water vapour (H_2O), and thus directly of OH radicals, whose reactions lead to the conversion of ozone-destroying NO and NO_2 catalysts to far less reactive nitric acid (HNO_3). On the other hand, reactions of OH radicals with hydrochloric acid (HCl), promote the formation of ozone-destroying Cl and ClO radicals. Methane is also a significant 'greenhouse gas': about 12 per cent of the added greenhouse warming forcing during the 1980s is estimated³ to come from CH_4 .

In the recent geological past, methane

would have caused a release rate from methane hydrates far too large to be consistent with the methane records.)

Consequently it is possible that 70 ± 15 Tg yr^{-1} of methane are released from leaks in the natural gas distribution systems and from natural gas and oil wells. If all of this comes from the gas energy sector, about 6–9 per cent of the natural gas production is leaked to the atmosphere. With such a high degree of leakage, not even a shift from coal to natural gas would alleviate future greenhouse warming. Some methane is also vented unburned from oil production sites, but the available statistics for the western world^{7–9} indicate that this cannot account for the unexplained 75 Tg yr^{-1} . This may point to a large degree of leakage from other parts of the world, or an unknown source of old methane. Clearly, this issue has to be resolved urgently in view of the forthcoming international global warming negotiations.

With about 100 Tg yr^{-1} of methane coming from the fossil-fuel sector, another 405 Tg yr^{-1} must come from biogenic sources. The best known among these are emissions from ruminant animals, about 80 Tg yr^{-1} . Landfills may release 50 ± 20 Tg yr^{-1} . Oceanic production and production by insects (termites) may contribute 30 Tg yr^{-1} .

ANDEAN MAGMA

Peeled or MASHed?

Rosemary Hickey-Vargas

How magma is generated at convergent tectonic plate margins and what the source materials are, have long puzzled geologists.

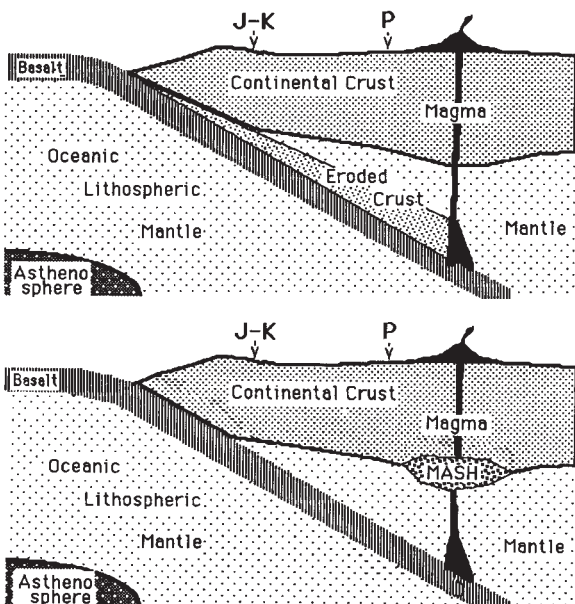


FIG. 1 Alternative explanations for the continental crustal signature in magmas from the northern SVZ: top, recycling of subduction eroded continental crust⁴; and below, contamination of magma at the lower crust/mantle boundary (MASH)⁵. Drawing modified from Stern¹. J–K and P indicate the position of Jurassic–Cretaceous and Pliocene plutonic belts, respectively.

Consequently, 245 ± 65 Tg yr^{-1} are released, mostly in the tropics, from natural wetlands, rice fields and biomass burning. Biomass burning may release 30 ± 15 Tg yr^{-1} (ref. 10), leaving an emission of 215 ± 50 Tg yr^{-1} from natural wetlands and rice fields. A stabilization of the atmospheric methane concentrations can be achieved through a 15–20 per cent reduction in the anthropogenic source³. It appears that tightening up emissions from landfills and the fossil fuel sector is the most practical way to achieve this goal. □

Paul J. Crutzen is at the Max Planck Institute for Chemistry, Postfach 3060, D-6500 Mainz, Germany.

1. Vaghjiani, G. L. & Ravishankara, A. R. *Nature* **350**, 406–409 (1991).
2. Levy, H. II. *Science* **173**, 141–143 (1971).
3. Houghton, J. T., Jenkins, G. J. & Ephraums, J. J. (eds) *Climate Change* (Cambridge University Press, 1990).
4. Pearnan, G. I. & Fraser, P. J. *Nature* **332**, 489–490 (1988).
5. Prinn, R. et al. *Science* **238**, 945–950 (1987).
6. Cicerone, R. & Oremland, R. *Global biogeochem. Cycles* **2**, 299–327 (1988).
7. Little, A. D. *Methane emissions from the Oil and Gas Production Industries* (Ruhrgas A. G., Essen, 1989).
8. *Ermittlung der Methan-Freisetzung durch Stoffverluste bei der Erdgasversorgung der Bundesrepublik Deutschland*, (Batelle Europe, Frankfurt am Main, 1989).
9. *Natural Gas Transmission and Distribution Methane Emissions Eng. Techn. Note SS-89-4-1* (Am. Gas Assoc., Washington DC, 1989).
10. Crutzen, P. J. & Andreae, M. O. *Science* **250**, 1669–1678 (1990).

Where subduction occurs along the margins of continents, such as the western coast of South America, there are diverse potential magma sources: the oceanic crust of the subducted tectonic plate (comprising basalt and marine sediment), an overlying wedge of the mantle and a cap of fusible continental crust can all contribute. Recent research on Andean magmatism has centred on the geochemical signatures of deep magma sources, particularly whether they are masked as the rising magma assimilates continental crust, and how far this is affected by crustal thickness. In contrast, Charles Stern¹ now suggests that continental crust may be assimilated not only during the late stages of Andean magma evolution, but that, in certain areas of the Andes, large amounts of crustal material may be introduced into the deep magma sources themselves by subduction erosion or scraping off of the continental margin.

Igneous rocks of the Andean mountain belt demonstrate much of what is known about subduction-related magmatism. For example, the chain of Andean

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

FIG. 2 Volcan Llaima, in central Chile, part of the southern SVZ.

volcanoes currently active, all situated approximately 100 km above the subducting oceanic plate and extending from southern Colombia to southernmost Chile, is broken into four segments, known as the northern volcanic zone (NVZ, southern Colombia and Ecuador), central volcanic zone (CVZ, southern Peru and northern Chile), southern volcanic zone (SVZ, central Chile) and austral volcanic zone (AVZ, southernmost Chile)^{2,3}. The segments are separated by 'volcanic gaps' lacking recent volcanic activity, which correspond to areas where the subducting tectonic plate is thrust under the continent at angles less than about 20°. This relationship shows that there must be a thick mantle wedge overlying the dehydrating subducted crust for magma to be generated. Similarly, the strongest geochemical indicators of assimilated continental crust are found in volcanic rocks from the CVZ and northern SVZ. In these areas, the continental crust is notably thick, greater than 70 km in the CVZ and at least 50 km in the northern SVZ; in the southern SVZ, by contrast, it is about 30 km thick.

The relationship between crustal thickness and geochemical signs of crustal contamination led to the concepts either that magma rising through thicker crust had more opportunity to interact or that the deeper portions of thick crust were more likely to melt and mix with rising magma⁴. This latter interpretation was promoted by Hildreth and Moorbath⁵ for the northern SVZ. Considering geochemical variations along the Andean chain within the SVZ, they envisage a thick, partially molten zone at the mantle/lower-crust interface, where magmas rising from deeper sources and partially molten lower crust undergo mixing, assimilation, storage and homogenization (MASH, for short).

But for the CVZ, Rogers and Hawkesworth⁶ pointed out that geochemical trends over the past 186 million years for igneous rocks on plots of $^{87}\text{Sr}/^{86}\text{Sr}$ versus Sr abundances and $^{143}\text{Nd}/^{144}\text{Nd}$ indicated two different magma source processes. One process (trend 1) they identified as the melting of Proterozoic mantle lithosphere within the mantle wedge, whereas the second (trend 2) involved assimilation of continental crust

during magma ascent. Younger volcanic rocks, which erupted after the thickening of the continental crust in this region, formed trend 2, thus strengthening the connection between crustal thickness and geochemical effects of assimilation.

Stern's argument for contamination at the source region by the subduction of crust eroded from the continental margin is based on geological evidence. In both the northern SVZ and the CVZ, belts of Jurassic (210–150-million-year-old) subduction-related plutonic rocks crop out near the coast, significantly closer to the subduction trench than younger, Tertiary plutonic rocks and the recent volcanoes. This implies a progressive trenchward migration of the older plutons, as the intervening crust was removed by subduction erosion. Stern estimates that erosion rates were about 80 km³ per million-years per km of trench for the northern SVZ and as much as 500 km³ per million years per km of trench for the CVZ, values that represent 1–7 per cent of the volume of the subducted plate. This quantity of recycled crust is sufficient to account for the observed crustal signature in CVZ and northern SVZ rocks.

Much of the geochemical argument for contamination of magma within the crust as opposed to contamination at depth is based on the behaviour of Sr, which is strongly partitioned into the mineral plagioclase. Plagioclase is stable in the crust to depths of about 30–50 km but is unstable in the mantle at similar depths, therefore direct melts of the continental crust have low Sr abundances (producing trend 2 of Rogers and Hawkesworth), whereas melts of the mantle, subducted crust and possibly the lowermost continental crust have high abundances (trend 1). Nevertheless, distinguishing source-region contamination from mantle heterogeneity or assimilation of plagioclase-poor lower crust is difficult.

One possible approach is to use along-strike and across-strike geochemical variations within segments of the Andes to determine the sequence in which geochemical trends are produced. For example, Morris *et al.*⁷ found that B–¹⁰Be systematics in SVZ volcanoes indicate mixing of two highly homogenous components along the entire length of this segment, including the northern SVZ. As one of these endmembers is the 'subduction component' (B and ¹⁰Be-rich), such homogeneity is not consistent

with mixing of varying proportions of subducted oceanic sediment and continental detritus along the arc, as implied by the source-contamination model.

Similarly, Tormey *et al.*⁸ present evidence that the geochemical effects of crustal assimilation in the SVZ are superimposed on variations produced by mantle composition and processes. Because Stern links a high rate of subduction erosion with a shallow subduction angle, an alternative approach might be to look for increasing evidence of crustal involvement in volcanoes of the southernmost SVZ near the intersection of the Chile Rise with the Chile Trench.

Although testing the subduction-erosion ARCHAEOLOGY

hypothesis may not be simple, it is clearly a viable alternative to the role of crustal thickness in producing the crustal signature in Andean magmas. An important perspective for research in this area is that many processes — upper and lower crustal assimilation, mantle heterogeneity and subduction erosion — are geologically plausible (if not certain) and none are mutually exclusive. The divergent models from different researchers may ultimately be part of a single complex story. □

Rosemary Hickey-Vargas is in the Department of Geology, Florida International University, Miami, Florida 33199, USA.

Ancient Britons in blue

Paul G. Bahn

FURTHER finds of fragments of human bodies have been made in the Cheshire peatbog which contained the 2,000-year-old remains of Lindow man. Strikingly, analysis of a piece of skin from the discoveries suggests that it had been painted¹. This constitutes the first archaeological evidence to support claims by Roman authors that the Ancient Britons stained themselves with some kind of pigment.

The first gruesome find made by the peatcutters of Lindow Moss, in Cheshire, UK, came to light in 1983 when a skull was recovered. Reported in the local press as that of a woman, it led a man who lived nearby to confess that he had murdered his wife some years before and buried her dismembered remains in the garden, which backed onto Lindow Moss. He should have kept silent, however, because the head, known as Lindow I, was subsequently shown to be at least 1,700 years old².

The principal find, Lindow II, followed in 1984, when the head and torso of a young, bearded man was recovered (see figure). Lindow II is probably the most thoroughly analysed and documented human body from the past², and has now been freeze-dried and

is exhibited in the British Museum in London. Radiocarbon dating has produced equivocal results which place him in the Iron Age or, more likely, the early Roman period.

In February 1987, another find was made, that of most of an adult body from the neck downward — this is Lindow III. Unfortunately, the remains suffered badly from the peatcutting process and were mashed into fragments. The body is of a male (the genitalia are present) who was probably in his early 20s, because the pelvic crest has not yet fused. One anomaly on a particularly well-preserved hand is the presence of an additional undeveloped digit just below the thumb. A large piece of gut has survived, which should permit further analyses of food and parasites like those carried out with such success on Lindow II. This body may in fact belong to the Lindow I head, which would therefore not be female after all.

A year later two human fragments — a left knee and a large piece of skin — were found, shortly followed by a piece of thigh. The skin comprises half of a right buttock, the whole of a left buttock, the anus and possibly some pubic hair. It is possible that these fragments, known as Lindow IV, belong to the lower



Lindow II: recovered from a Cheshire peat bog, and now the most analysed body from the past.

1. Stern, C. R. *Geology* **19**, 79–81 (1991).

2. Thorpe, R. S. *et al.* in *Andesites: Orogenic Andesites and Related Rocks* (ed. Thorpe, R. S.) 187–206 (Wiley, London, 1982).

3. Stern, C. R. *et al.* in *Andean Magmatism: Chemical and Isotopic constraints* (eds Harmon, R. S. & Barreiro, B. A.) 31–46 (Shiva, Cheshire, 1984).

4. Harmon, R. S. *et al.* *J. Geol. Soc. Lond.* **141**, 803–822 (1984).

5. Hildreth, W. & Moorbath, S. *Contrib. Mineral. Petrol.* **98**, 455–489 (1988).

6. Rogers, G. & Hawkesworth, C. J. *Earth planet. Sci. Lett.* **91**, 271–285 (1989).

7. Morris, J. D. *et al.* *Nature* **344**, 31–36 (1990).

8. Tormey, D. R. *et al.* in *Andean Magmatism and its Tectonic Setting* (eds Harmon, R. S. & Rapela, C.) (Geol. Soc. Am. Spec. Pap. in the press).

part of Lindow II. In other words, there were at least two bodies in this bog, though the actual total may be three or even four.

Preliminary investigation of Lindow III has revealed no evidence of necrophilous elements such as fly pupae or the beetles which feed on them, which implies that the body was not thrown into an open pool but rather was deliberately pushed under the peaty mud immediately after death. But most intriguing are the results of electron probe X-ray microanalysis of a small piece of skin from one shoulder carried out by Pyatt *et al.*¹ — unlike Lindow II's skin, that of Lindow III contains an extremely high level of copper ions together with an abnormally high amount of aluminium.

The investigators believe that these elements cannot be explained as components of normal skin tissue or as the consequence of ions moving in from the surrounding peat. Instead they think it probable that a foreign substance had been applied to the skin surface, and penetrated through the pores. The copper, in particular, suggests a pigment of some kind. Similar analysis of Lindow II was restricted to a search for vegetable substances rather than mineral ones, but a green fluorescence on the fox-hair band around his arm might be the result of copper

ECHOLOCATION

enrichment.

Written records of painted Ancient Britons start with Caesar, who mentioned that they "stain themselves with *vitrum*, which gives a blue colour and a wilder appearance in battle". Pliny specified that a vegetable dye was used, "from a plant like a plantain called *glastum* in Gaul" — this has usually been translated as woad, though that plant (*Isatis tinctoria*) bears no resemblance to a plantain. Caesar's word *vitrum*, which normally means glass or crystal, has also been translated as woad since the sixteenth century, when the plant was a popular source of blue dye; but the oldest known occurrences of the woad plant in Britain are from the Anglo-Saxon period.

Pyatt *et al.* therefore think that *vitrum* probably refers not to a plant but to a copper-based pigment of the kind they have detected. However be that may, the fearsome blue figure of the Ancient Briton now takes on fresh credibility. □

Paul G. Bahn, 428 Anlaby Road, Hull HU3 6QP, UK, is a freelance writer on archaeology.

1. Pyatt, F. B., Beaumont, E. H., Lacy, D., Magilton, J.R. & Buckland, P.C. *Oxford J. Archaeol.* **10**, 61–73 (1991).
2. Stead, I. M., Bourke, J. B. & Brothwell, D. (eds) *Lindow Man: The Body in the Bog* (British Museum, London, 1986).

The cost of being a bat

Jeremy M. V. Rayner

THE energetic expense of being a flying and echolocating bat is surprisingly modest, according to the report by Speakman and Racey which appears on page 421 of this issue¹. The reason is the happy way in which echolocation, respiration and flight in bats are coupled.

Two years ago, Speakman, Anderson and Racey² described measurements of echolocation in resting pipistrelles (*Pipistrellus pipistrellus*) and came to the conclusion that the cost of echolocation is of the same magnitude as the large cost of flight itself. But are the energy expenditures involved in flight and in echolocation additive? To answer the question, Speakman and Racey used the same techniques to determine the energy cost of flight in the pipistrelle, and also in the long-eared bat *Plecotus auritus*. These experiments are valuable because measurements of flight costs in microchiropteran bats are scarce; moreover, they combine two commonly used techniques — respirometry and isotopic measurement of labelled water metabolism — and therefore validate both methods, and avoid some of the pitfalls of either in isolation.

The measured flight energy rate of 1.3 W in the pipistrelle is greater than the estimated 0.9 W for echolocation alone², but not by as much as if the costs of the two activities were simply added. To confirm that finding, Speakman and Racey determined how the

energy cost of flight in echolocating microbats varies with size, and compared the pattern with that for flight in birds and in non-echolocating megachiropteran bats; there was no significant difference between the two groups. The microbat sample is small (the new measurements double its size), and many of the bird measurements were made in wind tunnels, with all the errors of that methodology, so the close agreement is particularly striking. Echolocation while not in flight might be expensive — although, at rest, call pulse rates are low to compensate — but echolocation in flight apparently adds little or nothing to the total energetic cost.

What is the physiological and biomechanical background to this coupling between locomotion and respiration? Terrestrial mammals tend to synchronize breaths with strides³ — the stride:breath ratio becomes tighter and closer to 1:1 as the animal moves faster, and the energy cost, the forces generated in locomotion and the oxygen demand all increase. This coupling is absent in running lizards, because of their stance and running gait and the 'slow' physiology of the relevant thoracic muscle⁴. Bramble⁵ proposed a 'visceral piston' mechanism by which displacements of the gut of a fast-moving animal alternately expand and compress the lungs, driving respiration, but Alexander⁶ has shown that the visceral piston cannot explain the observed phasing of respiration

Copying connection

DOWN'S syndrome, the most common genetic form of mental retardation, is caused by an additional copy of all or part of chromosome 21. But from which parent does the extra copy arise? Results of a study carried out by S. E. Antonarakis *et al.* (*New Engl. J. Med.* **324**, 872–876; 1991), in which 200 Down's syndrome children were examined with 20 polymorphic DNA markers from the long arm of chromosome 21, reveal that it is maternal in origin over 95 per cent of the time; this proportion is significantly higher than had been estimated from previous studies. The predominance of maternal non-disjunction in the pathogenesis of the syndrome supports the notion that most cases are caused by a defect during oocyte meiosis.

Breathless

SANDS deposited in the Witwatersrand basin in South Africa confirm that the atmosphere when the Earth was young had an extremely low concentration of oxygen, argue J. B. Maynard and colleagues (*Geology* **19**, 265–268; 1991). The claim is made because the sand, dating back to 3,000 million years ago, contains samples of uranite, a mineral that is easily oxidized. The idea had been challenged, because uranite has also been found in young sands in the Indus river. But Maynard *et al.* show that the South African sand has been exposed to extremely severe weathering, which would have permitted an oxygen-rich atmosphere to attack the uranite, whereas the Indian sand is in a relatively pristine condition.

Ends and means

ANY number of actin-capping proteins are known — such proteins bind to the 'barbed' (fast) end of the filament, after which all action at this end ceases and growth or shortening are confined to the other ('pointed') end. But one of them, insertin, behaves differently, for although it binds strongly it allows the barbed end to maintain its equilibrium with the monomer pool and growth remains possible. A. Gaertner and A. Wegner (*J. Muscle Res. Cell Motil.* **12**, 27–36; 1991) have investigated how this comes about, and from an analysis of the kinetics they have been able to exclude any model in which the insertin hops off to allow attachment of a new terminal actin subunit and then rebinds. All of the evidence points to a process whereby the terminal subunit can slide in or out under the bound integrin. This is reminiscent of the growth of microtubules from the captive end on a kinetochore and constitutes a new mechanism of length change, allowing perhaps elongation from the plasma-membrane-bound filament end in the cell.

Ardea



Upbeat — the coupling between respiration and locomotion in the dog dictates that the bark synchronizes with stride and is emitted when the forelimbs come to the ground.

in dogs or horses (although it might apply to hopping wallabies). He proposed two alternative mechanisms which might explain the coupling, namely bending of the back, and loading of the fore-limbs on the thorax; the first of these seemed the most appropriate explanation for the horse.

In flight the situation is rather different. It is unlikely that the visceral piston or back-bending mechanisms might drive respiration because both are axial movements of the body, whereas the main movements of the body and wings in flight are lateral. Some authors have argued that there is little or no coupling between respiration and flight in birds, and that such coupling cannot be expected owing to the through-flow structure of the bird lung and the anatomy of the avian chest skeleton⁶. The experimental evidence is equivocal — although some observations have shown very poor coupling between inhalation and wingbeat (usually for very short anaerobic flights), the few observations of sustained flights show that coupling is tightly 3:1 or 2:1, or in extreme exertion even 1:1 (ref. 7). It is known that deformation of the thorax and pectoral girdle under the large mechanical loadings from the wings in flight can drive air flows into and out of the air sacs⁸. In other words, the third of Alexander's three mechanisms could apply to birds.

The aerodynamics of bird and bat flight are very similar⁹, and the thorax will deform under comparable loads from flapping wings which, as in birds, could drive phase locking between the wingbeat and respiration. In fact the argument comes full circle, and it is echolocation which provides the best evidence of this coupling. Flying microbats emit echolocation pulses once on each wingbeat, invariably starting in the latter phase of the upstroke, coinciding with exhalation^{10,11}. At first sight this would appear to be the wrong phase of the wingbeat, because aerodynamic

lift is greatest in mid-downstroke, so this should be the phase when the thorax is compressed by the flight muscles¹¹. However, as usual, things are not so simple because forces at the wing root are dominated by wing inertia, not by lift: the greatest force to pull the wing downwards occurs at the end of the upstroke⁹. It is at this phase that air is expelled from the lungs, and the bat emits its echolocation pulse 'on the back' of breathing out.

This is the mechanism by which echolocation, respiration and flight in microbats are related, and it explains why echolocation in flight is relatively economic. The bat must respire to obtain oxygen for flight, and can use the exhaled air pulses to carry its ultrasonic cries with minimal extra physiological or energetic cost. This also fits with the scarcity of echolocation in non-flying tetrapods: exhaled airflows are not sufficiently concentrated to carry enough sound energy to make echolocation worthwhile. Echolocation is not found in most birds because — in comparison with vision — it is of little use to a daytime flier⁴; there is much still to tell about its absence in birds such as owls and nightjars.

More puzzling is the lack of ultrasonic echolocation in the megabats, where it is confined to a crude form in the cave-dwelling *Rousettus*. As nocturnal fliers, megabats would undoubtedly benefit from it for orientation, even if it cannot help much in finding fruit or nectar. In this context the absence (or loss?) of echolocation in this group is surprising: microbats have a much older fossil record, and the earliest known Palaeocene microbats were already echolocating aerial insectivores. This apparent discrepancy is however fully consistent with the belief that bats are diphyletic¹², that megabats evolved from primates sometime after microbats had become established; primates already possessed well-developed vision, and the evolutionary pressures surrounding the development of echolocation in the visually acute proto-megabats may not have been sufficient to favour the extreme specialization it would have involved. □

Jeremy M. V. Rayner is in the Department of Zoology, University of Bristol, Woodland Road, Bristol BS8 1UG, UK.

1. Speakman, J. R. & Racey, P. A. *Nature* **350**, 421–423 (1991).
2. Speakman, J. R., Anderson, M. E. & Racey, P. A. *J. comp. Physiol.* **A165**, 679–685 (1989).
3. Bramble, D. M. *Am. Zool.* **29**, 171–186 (1989).
4. Carrier, D. R. *J. exp. Biol.* **143**, 435–457 (1989).
5. Alexander, R. McN. *J. Zool., Lond.* **218**, 69–85 (1989).
6. Duncker, H.-R. in *Complex Organismal Functions: Integration and Evolution in Vertebrates* (eds Wake, D. B. & Roth, G.) 147–169 (Wiley, New York, 1989).
7. Butler, P. J. & Woakes, A. J. in *Bird Migration* (ed. Gwinner, E.) 300–318 (Springer, Heidelberg, 1990).
8. Jenkins, F. A., Dial, K. P. & Goslow, G. E. *Science* **241**, 1495–1498 (1988).
9. Rayner, J. M. V. in *Recent Advances in the Study of Bats* (eds Fenton, M. B., Racey, P. A. & Rayner, J. M. V.) 23–42 (Cambridge University Press, 1987).
10. Schnitzler, H.-U. *Z. vergl. Physiol.* **73**, 209–221 (1971).
11. Thomas, S. P. in *Recent Advances in the Study of Bats* (eds Fenton, M. B., Racey, P. A. & Rayner, J. M. V.) 75–99 (Cambridge University Press, 1987).
12. Pettigrew, J. B. et al. *Phil. Trans. R. Soc. Lond.* **B325**, 489–559 (1989).

Diamond mill

THE distant dream of controlled nuclear fusion continues to absorb billions of pounds of taxpayers' money. Laser fusion is probably the current front runner. This requires a tiny fuel pellet containing tritium and deuterium to be dropped through an optical focus on which a number of powerful pulsed lasers are aimed. As the pellet reaches the focal point, the lasers all fire at once, hitting the pellet from all directions with hundreds of kilojoules of energy packed into a few nanoseconds. Its surface is vaporized. Such is the inertia of material on this short time scale that the expanding vapour cannot simply boil away. It squeezes and heats the inside of the pellet to nuclear fusion point, releasing a sudden flare of energy. In a working power station, fuel pellets will be dropped into the fusion chamber as fast as the lasers can recycle — maybe twenty times a second. Gigawatts of power will result.

Well, that's the dream, the crock of gold at the laser's end. Daedalus now proposes that the vastly expensive fusion laboratories should earn some honest money in the meantime. They study other matters as well, of course; this journal recently carried a letter from one of them concerning the transformation of graphite into diamond (Erskine, D. J. and Nellis, W. J. *Nature* **349**, 317; 1991). It seems that this is a fast, martensitic process. It takes only ten nanoseconds, just about the length of a fusion-laser pulse, and needs — a mere 300,000 atmospheres — far below what can be achieved by laser-compression. So Daedalus now suggests that the laser-fusion workers abandon their silly little pellets of deuterium-tritium mixture, and start dropping spheres of graphite into their laser-focus.

Crude calculations suggest that existing inertial-confinement laser systems might well be able to transform a 20-gram graphite pellet into a 50-carat diamond at every pulse. At an output rate of 20 Hz, even allowing for the expense of cutting and polishing the gems, this could bring in well over a million pounds a second. If used to generate fusion electricity, the same equipment would be lucky to bring in a million pounds a week.

Fusion lasers on the drawing board are ten times more powerful still, making possible the mass-production of diamonds up to 500 carats or so. World prices will undoubtedly tumble, and de Beers will be apoplectic, but even so the operation should turn in a tidy profit. Mass-produced bulk diamonds will be wonderfully useful for spectroscopic windows and prisms, heat-sinks, cutting-tools, chemical flasks and crucibles, and cheap jewellery of all kinds. Elegant socialites will attend glittering receptions in their diamond tiaras, while leaving the real paste ones carefully locked in the safe.

David Jones

Warming in the Arctic

SIR — Climate models predict that global warming caused by the greenhouse effect will not be distributed evenly¹. The greatest temperature increase is expected for the polar and subpolar regions, with a relatively small change in the tropics. The polar regions may therefore be well suited to attempts to observe climate changes, but severe environmental conditions make it difficult to obtain such data, particularly in the case of the Arctic Ocean, which has a permanent ice cover and is thus virtually inaccessible to normal ocean-going vessels.

through turbulent mixing and diffusive fluxes.

Climatological² and synoptic data collected in 1987 (ref. 3) show that the maximum temperature in the Atlantic water decreased from about 4 °C in the ice-free Fram Strait to below 2 °C in the subsurface layer between Svalbard and Severnaya Zemlya. But our temperature measurements made in 1990 off Frans Joseph Land and north of Severnaya Zemlya had maxima of 2.8 °C, about a degree higher.

The ice thickness encountered during the

monitoring of the heat fluctuations in the Arctic, a region so vulnerable to changes in global climate.

DETLEF QUADFASSEL

Institut für Meereskunde,
2 Hamburg 54, Germany

ALEXANDER SY

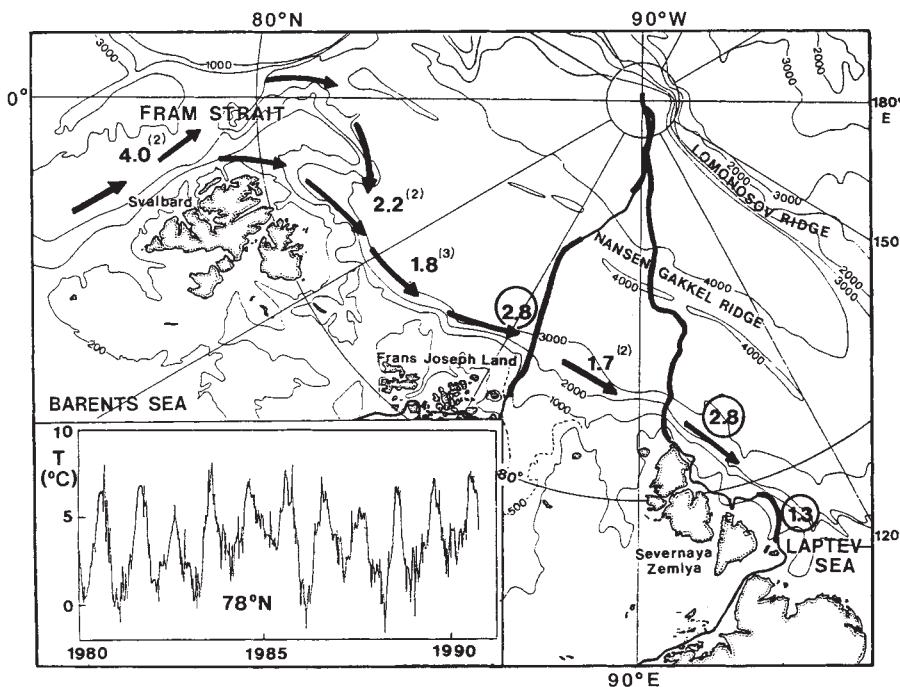
Bundesanstalt für Seeschifffahrt
und Hydrographie,
2 Hamburg 36, Germany

DOUGALD WELLS

Norland Science and Engineering Ltd,
Ottawa, Ontario K1P 5G8, Canada

ALFRED TUNIK

ABS Group, Paramus,
New Jersey 07653, USA



Maximum temperature in the Atlantic layer over the Eurasian Basin of the Arctic Ocean. Circled numbers, *Rossiya* observations along the tracks shown as bold lines. Other numbers are taken from the references given in brackets. The inset shows the maximum sea surface temperature in Fram Strait at 78°N during 1980–90.

During August 1990, the Soviet ice-breaker *Rossiya* carried a group of tourists from Murmansk to the North Pole, cutting across the Eurasian basin at longitudes 60° E and 95° E. Detailed observations of the ice were made as well as closely spaced temperature profile measurements in the upper 500 m of the water column, using expendable bathythermographs dropped in leads or small polynyas.

The Arctic Ocean water column consists principally of three layers: a cold and low salinity surface layer; a warm and salty Atlantic layer in the depth range 100–800 m; and the deep water below, being cooler and slightly fresher. The Atlantic layer derives from an inflow in the Fram Strait between Svalbard and Greenland and is the northernmost extension of the Gulf Stream–Atlantic current system. On its way north it loses heat to the atmosphere, and in or just north of Fram Strait subducts underneath the ice and polar water, continuing to lose its heat

cruise was 20–30% smaller than the climatological mean for the area⁴, although this might simply reflect that the vessel was taking the route of least resistance, avoiding thick old massives and other heavy ice features. It is of course not immediately clear whether the higher subsurface temperatures and the lower ice thickness already reflect a climate trend or just some normal interannual variability. The sparseness of the available data simply does not allow any such conclusion. Sea-surface temperature variability in Fram Strait is high, both on seasonal and interannual timescales. The year-to-year changes of 2 °C could mask any trends in temperature development. It seems clear, however, that there is a need for careful

Clonal defence

SIR — Dye *et al.*¹ have criticized our clonal theory of parasitic protozoa² by pointing out that “clonality is the conventional wisdom” and that “sexual reproduction clearly occurs, [although . . .] may not be common in nature”. Unfortunately, they fail to understand the significance of our theory.

We are not denying that sexual reproduction may occur nor simply asserting that clonality may be a common mode of reproduction for parasitic protozoa, both of which are indeed well known. Our hypothesis is that parasitic protozoa have a clonal population structure, which in turn implies that sexual reproduction is very rare even on the evolutionary scale. The theory has momentous evolutionary as well as medical consequences. In *Trypanosoma cruzi*, for example, where the evidence supporting the hypothesis is overwhelming, the average genetic differentiation between two randomly sampled clones is greater than that between gorillas and humans. Generally, if the population structure is clonal it is not warranted to assume that any clone or a few of them may be representative of the species with respect to biological and pathological properties.

Walliker *et al.*³ argue that our theory is likely to be wrong with respect to the agent of human malaria *Plasmodium falciparum*. This may be so, but none of their four points will settle the issue. Because of the well-known fact that a sexual stage is required in the mosquito vector to complete the infection cycle, we were surprised that our analysis² of the limited data available supported a clonal population structure. In addition to a clonal population structure for the whole taxon, it is also possible that both biparental and uniparental lineages may coexist in nature — a phenomenon known to occur in many metazoa and plants.

The issue of the population structure of *P. falciparum*, as in the other protozoa, can be settled only by population-genetic evidence concerning the distribution of genotypes in natural populations. We reported the analysis supporting clonality as a challenge to those who work on *P. falciparum* to settle the

1. Mikolejczak, U., Santer, B. D. & Maier-Reimer, E. *Nature* **345**, 589–593 (1990).
2. Gorshkov, S. G. *Atlas of the Oceans: The Arctic Ocean* (Pergamon, New York, 1983).
3. Anderson, L. G. *et al.* *Deep Sea Res.* **36**, 475–482 (1989).
4. Bourke, R. H. & Garret, R. P. *Cold Reg. Sci. Tech.* **13**, 259–280 (1987).

issue, given its great biological and medical significance. We agree with Walliker *et al.*³ that allelic frequencies and the number of loci must be taken into account for estimating the probability of observed multilocus genotypes, and did this as best we could based on the scarce published evidence and obtained statistically significant evidence of nonrandom association between loci (Table 2 of ref. 2.).

Dye *et al.*¹ suggest that "analyses of protozoan clonality, particularly *Plasmodium*, will yield more convincing results if carried out on restricted geographic scales... not by making intercontinental comparisons...". We agree that more population-genetic data are needed: we could only analyse such data as have been published. But Dye *et al.* do not seem aware that the most convincing evidence for clonal population structure would come precisely from the observation of identical multilocus genotypes in geographically distant localities. Moreover, if such widespread genotypes exist, as seems to be the case in *Trypanosoma cruzi*,

for example, they are the ones that should most urgently become the subject of biological and medical research.

MICHEL TIBAYRENC

Laboratoire de Genetique des
Parasites,
ORSTOM, BP 5045,
34032 Montpellier Cédex,
France

FINN KJELLBERG

CEPE, CNRS,
Route de Mende,
34033 Montpellier Cédex,
France

FRANCISCO J. AYALA

Department of Ecology and
Evolutionary Biology,
University of California,
Irvine, California 92717,
USA

1. Dye, C., Davies, C. R. & Lines, J. D. *Nature* **348**, 120 (1990).
2. Tibayrenc, M., Kjellberg, F. & Ayala, F. J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2414–2418 (1990).
3. Walliker, D., Beale, G. & Luzzatto, L. *Nature* **348**, 120 (1990).

Sea-level rise and earthquakes

SIR — Global sea level is believed to have risen at a mean rate of 1.7 ± 0.4 mm per year during the past 80 years¹. Several models suggest that sea-level rise is accelerating². To be of benefit to coastal communities, useful forecasts of sea-level rise can be made only by estimating the rise over decade or shorter timescales, which will require significant improvements in our misunderstanding of global ocean dynamics and much improved sea-level sampling of mid-ocean and Southern Hemisphere locations³.

As our ability to measure global sea level improves, it will become necessary to suppress the solid Earth input to global sea level, hitherto assumed to be an insignificant but steady contribution to sea-level change at decadal periods⁴. Although the net volume of the ocean basins increases slightly because the Alpine/Himalayan tectonic convergence zones are continental collisions, for example, the resulting reduction in global sea level is inferred to be less than 10 μ m per year. However, the sea-level perturbation from great oceanic subduction earthquakes is both significant and rapid. The Chilean earthquake in 1960 was accompanied by sea floor uplift and subsidence⁵ with a net reduction in ocean basin volume of more than 573 km³. This local reduction in ocean basin capacity causes an increase in global sea level of 1.7 mm, equal to the annual mean global sea-level rise from other causes. Similarly, the 1964 Alaskan earthquake⁶ was accompanied by a reduction in ocean basin capacity of approximately 250 km³ due to uplift, resulting in a 0.7-mm increase in global sea level. Offsets in tide gauge records following these events are obscured by daily and weekly oceanographic noise.

It is clearly important to remove the effects of submarine deformation accompanying subduction processes from future estimates of trends in sea level. This will require improved estimates of submarine co-seismic deformation, particularly in the far field and in the interseismic period. For example, the net long-term contribution to global sea level from subduction earthquakes is probably close to zero given reasonable models for post-seismic deformation and interseismic strain development. Thus global sea level must subside between great earthquakes. Because offshore postseismic and interseismic deformation data are unavailable for most tsunamigenic⁷ earthquakes, we are surprisingly ignorant of the true magnitude of this subsidence. Making the simplifying assumptions that coseismic sea-level uplift scales linearly with magnitude, and that interseismic deformation is linear in time between events, we infer a minimum sea-level lowering rate between historical earthquakes with magnitudes $M_w > 8.5$ of approximately 0.05 mm per year.

ROGER BILHAM

Department of Geological Sciences,
University of Colorado,
Boulder,
Colorado 80309,
USA

SERGIO BARRIENTOS

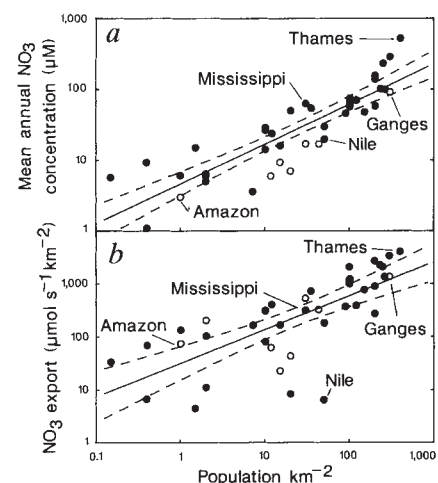
Department of Geology and
Geophysics,
University of Chile,
Santiago, Chile

1. Trupin, A. & Wahr, J. *Geophys. J. Int.* **100**, 441–453 (1990).
2. Meier, M. *Nature* **343**, 115–116 (1990).
3. Bilham, R. *Rev. Geophysics* **29**, 1–29, (1991).
4. Harrison, C. G. A. *Eos, Trans. Am. geophys. Un.* **70**, 1002 (1989).
5. Pfaffner, G. & Savage, J. *Geol. Soc. Am. Bull.* **81**, 1001–1030 (1970).
6. Pfaffner, G. *Science* **148**, 1675–1689 (1965).
7. Abe, K. *J. geophys. Res.* **84**, 1561–1568 (1979).

Human influence on river nitrogen

SIR — The overabundance of algae in coastal waters is often blamed on excessive inputs of otherwise scarce nutrients such as nitrate and other forms of nitrogen¹. This nutrient overload originates primarily from rivers that discharge directly into an ocean^{2,3}. A key question, therefore, is what controls the amount of nutrients in, and exported from, coastal rivers? We now have quantitative evidence that, on a global scale, human population within a river's watershed is strongly related to the concentration of nitrate in rivers that discharge to coastal ecosystems.

We collected published data for 42 major rivers of the world (references available on



a, Average annual nitrate concentration in 42 rivers of the world against the corresponding human population density in the rivers' watersheds. Each point represents a different river; some major rivers have been labelled for comparison. ○, tropical rivers ($\leq 23.5^\circ$ lat.); ●, all other rivers. Solid line, least squares regression: $\log[\text{NO}_3] = 0.56 \times \log[\text{Pop. Dens.}] + 0.67$ ($r^2 = 0.76$, $P < 0.00001$); dashed lines, 95% confidence intervals for the regression. b, Same as a, but for nitrate export. The regression line is $\log[\text{NO}_3 \text{ Export}] = 0.64 \times \log[\text{Pop. Dens.}] + 1.51$ ($r^2 = 0.53$, $P < 0.00001$).

request). The rivers chosen were restricted to those that charge into an ocean and that had the appropriate data available — watershed population density, watershed area and mean annual nitrate concentration and water flow near the river mouth. The selected rivers appear to be representative of the world's rivers; mean runoff (flow per watershed area) is 13.2 litres s⁻¹ km⁻² compared to a world mean value of 11.8 litres s⁻¹ km⁻² (ref. 4). The rivers studied and their watersheds account for about 37% of global freshwater discharge to the ocean⁴.

A log-log regression of mean annual nitrate concentration plotted against human population density (part a of figure) shows a highly significant relationship ($P < 0.00001$) that explains 76% of the variation

in nitrate concentration for the 42 rivers. We repeated the analysis for a calculated nitrate export (concentration $\times$ mean runoff) and found a similar significant relationship ($P < 0.00001$) that accounts for 53% of the variation in nitrate exported to coastal waters (part *b* of figure). Other watershed characteristics such as area and water flow showed, in general, no significant relationship to either nitrate concentration or nitrate export.

River nitrate concentration and export are affected by complex biotic, abiotic and anthropogenic factors, so it is striking that they show such a marked relationship to population density. Other variables, such as water flow and watershed area, did not explain a substantial amount of variation when included in multiple regression analyses. Our study demonstrates that human population density in a watershed is the best predictor of river nitrate concentrations and export.

If we consider human influences alone, river nitrate could be affected by sewage loading, atmospheric deposition, agriculture and deforestation. We used an estimate of per-capita nitrogen loading in sewage⁵ and found that sewage loading was the same magnitude as nitrate exported from the selected rivers. We also found a significant relationship between calculated atmospheric nitrate deposition and river nitrate concentration, confirming studies that show atmospheric deposition to be a significant source of nitrogen in rivers^{3,6}. We did not quantify the impact of agriculture and deforestation, but these additional human disturbances have been shown to contribute significantly to nitrate exported from watersheds^{7,9}.

Human activity clearly dominates nitrate export from land, whether or not we take into account all sources of anthropogenic nitrate. Projected global population growth will potentially result, therefore, in increased nitrate export from rivers, with often undesirable consequences for coastal ecosystems.

BENJAMIN L. PEIERLS
NINA F. CARACO
MICHAEL L. PACE
JONATHAN J. COLE

*Institute of Ecosystem Studies,
The New York Botanical
Garden,
Mary Flagler Cary Arboretum,
Box AB, Millbrook,
New York 12545, USA*

1. D'Elia, C. F. *Environment* **29**, 6–11; 30–33 (1987).
2. Walsh, J. J., Rowe, G. T., Iverson, R. L. & McRoy, C. P. *Nature* **291**, 196–201 (1981).
3. Fisher, D. C. & Oppenheimer, M. *Ambio* (in the press).
4. Meybeck M. *Am. J. Sci.* **282**, 401–450 (1982).
5. Vollenweider, R. A. *OECD Tech. Report DAS/CSI/68.27* (1968).
6. Paerl, H. W. *Nature* **315**, 747–749 (1985).
7. Likens, G. E., Bormann, F. H., Johnson, N. M., Fisher, D. W. & Pierce, R. S. *Ecol. Monogr.* **40**, 23–47 (1970).
8. Kohl, D. H., Shearer, G. B. & Compton, B. *Science* **174**, 1331–1334 (1974).
9. Haycock, N. & Burt, T. *Nature* **348**, 291 (1990).

DNA fingerprinting in India

SIR — In India, the DNA fingerprinting test was used for the first time during June 1989 to settle a drawn-out paternity case in Madras^{1,2}. Another court judgement based on DNA evidence was given in a dispute between an unmarried mother and her lover in Kerala state^{2,3}. This biotechnology-based commercial service is mainly carried out by a team at the Centre for Cellular and Molecular Biology, Hyderabad. The *Bkm* (banded krait minor satellite DNA) probe developed by this team costs one-tenth of the tests used in Europe or the United States. Paternity-disputed court cases are much more common in India, and most are now being referred to the Centre for Cellular and Molecular Biology for DNA evidence.

Several reports^{4–7} have raised doubts about the accuracy of these tests in identifying suspects in criminal cases. The question raised by the population geneticists is of serious concern to this new forensic tool. There are special problems in using the tests in India, a tropical country with heterogeneous environment, covering a latitudinal range of about 28°. Environmental factors change the overall genetic structure of the populations that regularly change with latitude. This phenomenon exists in *Drosophila melanogaster* chromosome inversion polymorphism (A. D. and B. N. Singh, manuscript in preparation). Different human subpopulations exist in India (with respect to religion, caste and subcaste), each with a different history of settlement and migration pattern. Castism is acute, and intracaste marriages are still predominant over the intercaste marriages. Thus, there are various subpopulations of random mating in which genetic material is shuffled within a gene pool that remains isolated from other such populations. Therefore, it is much more difficult to prepare a common population genetic database in India than in other countries because genotype frequency, or the frequency of a particular pair of alleles at a given locus, cannot be calculated — thus it is impossible to calculate the Hardy–Weinberg equilibrium for the whole population. Also, it cannot be assumed that the allele types are shuffled at random, or that the occurrence of one allele is dependent on the second allele. It is possible that in some populations, alleles are tightly linked to others present either intra- or interchromosomally.

A low-cost DNA evidence test would be a step forward for a developing country like India, but the question lies in reliability. Countries such as the United Kingdom and the United States are still in search of clear, common guidelines for the DNA profiling test and a correct genetic database for populations to interpret the observed results. In India, with its different subpopulations, it is an uphill task to construct a correct genetic database. The problem could be solved by

preparing separate databases for different randomly mating subpopulations. Separate genotype frequencies could be obtained and calculation of linkage equilibrium between loci could be performed for such populations. The databases thus prepared could be used to estimate the rarity of a common allele in the given population in which the DNA match is to be performed.

There should be special legislation to cover the results of DNA analysis, but there is at present no such legislation on the admissibility of DNA evidence in India. As the opinion of the expert is valid in court under Indian law³, such 'experts' should be very careful in interpreting results.

APARUP DAS

*Genetics Laboratory,
Centre of Advanced Study in Zoology,
Banaras Hindu University,
Varanasi-221 005, India*

1. Jayaraman, K. S. *Nature* **340**, 175 (1989).
2. Das, A. *Inven. Intell* (in the press).
3. Jayaraman, K. S. *Nature* **346**, 208 (1990).
4. Lander, E. S. *Nature* **339**, 501–505 (1989).
5. Neufeld, P. J. & Colman, N. *Scient. Am.* **262**, 46–53 (1990).
6. Joyce C. *New. Scient.* **127**, 24–25 (1990).
7. Thompson, W. C. & Ford, S. *New. Scient.* **125**, 38 (1990).

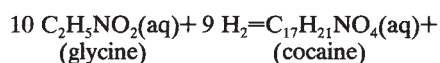
Extraterrestrial synthesis

SIR — Shock and Schulte¹ have calculated the metastable equilibrium concentrations of amino acids and other organic compounds formed from pyrene, fluoranthene, CO₂, NH₃ and H₂O. They propose that this was the mechanism of synthesis of the amino acids on the Murchison meteorite parent body. The pyrene and fluoranthene are thought to have been produced in carbon stars or on interplanetary dust. This calculation is accepted by others² as a mechanism for the synthesis of some Murchison organic compounds.

But there are fundamental errors in these types of thermodynamic calculations. Organic reactions that involve the breaking and reforming of many carbon–carbon bonds generally do not come to equilibrium at temperatures below about 700 °C. Dehydrogenations and ester/ether formation come to quasi-equilibrium at lower temperatures, but no carbon–carbon bonds are broken in these reactions. At higher temperatures equilibrium may be obtained, but the equilibrium concentrations of organic compounds that contain more than a few carbons are extremely low. Catalysts able to achieve such favourable equilibria at low temperatures are not known.

Shock and Schulte's choice of pyrene and fluoranthene as starting materials is arbitrary; different results would be obtained with other, more abundant organic molecules found in the Murchison meteorite. Their choice of products is also arbitrary, as many other compounds are possible: for example there are 551 isomers of decanoic acid alone³. An arbitrary choice of reactants

and products can lead to ridiculous results. For example, consider the synthesis of cocaine from glycine:



Using the free energies of formation $\Delta G_f^\circ = -88.58 \pm 0.16 \text{ kcal mol}^{-1}$ for glycine⁴ and $\Delta G_f^\circ = -65.8 \pm 10.0 \text{ kcal mol}^{-1}$ for cocaine (our estimate) and the standard ΔG_f° for NH_3 , CO_2 and H_2O (refs 4,5) gives $\Delta G_{\text{react}}^\circ = -64.0 \pm 10.0 \text{ kcal}$ at 25 °C and an equilibrium constant $K_{\text{eq}} = 10^{47}$. Using this equilibrium constant, and taking NH_3 fugacity = 10^{-3} atm and CO_2 fugacity = 25 atm, as specified by Shock and Schulte (even though these values are unrealistic), we calculate the cocaine to glycine ratio as a function of the H_2 fugacity, f_{H_2} . Taking glycine as $4 \times 10^{-3} \text{ M}$, which would be the concentration if 10 p.p.m. glycine were dissolved in the interstitial water (10% by volume) of the Murchison parent body⁶, we find that the cocaine and glycine concentrations will be equal at $f_{\text{H}_2} = 4 \times 10^{-6} \text{ atm}$ ($\log f_{\text{H}_2} = -5.4 \pm 0.8$). A cocaine concentration of $4 \times 10^{-3} \text{ M}$ is close to saturation ($6 \times 10^{-3} \text{ M}$) (ref. 7), and so cocaine would precipitate out at slightly higher f_{H_2} values. At $f_{\text{H}_2} = 7 \times 10^{-4} \text{ atm}$, 99% of the glycine would be converted to cocaine. These f_{H_2} values are in agreement with those calculated from the estimated redox potentials during aqueous alteration on the carbonaceous chondrite parent body⁸. As far as we know, no meteorites contain profuse amounts of cocaine.

Although the assumption of thermodynamic equilibrium generally leads to a good description of inorganic reactions, thermodynamic equilibrium is attained only rarely with organic compounds. Calculations based on thermodynamic equilibrium can generate misleading conclusions for organic mixtures.

STANLEY L. MILLER

Department of Chemistry,

JEFFREY L. BADA

Scripps Institution of Oceanography,
University of California, San Diego,
La Jolla, California 92093, USA

SHOCK AND SCHULTE REPLY — Natural systems exist in states of partial equilibrium in which some species are preserved far from equilibrium by kinetic barriers and others coexist at stable or metastable equilibrium. To identify metastable equilibrium states, realistic physical and chemical parameters that prevail in natural systems must be evaluated. Evidence from nature, elucidated with thermodynamic calculations, indicates that over millions of years the complex mixtures of organic compounds in sedimentary basins approach metastable equilibrium states at temperatures as low as 100 °C (ref. 9). At low temperatures, the enormous lengths of time required to reach metastable equilibrium states make laboratory experiments

impossible. But such states may be reached on laboratory time scales at higher temperatures, as demonstrated in our reinterpretation¹⁰ of experiments conducted by Miller and Bada¹¹, in which a metastable state among aqueous amino acids was established at 250 °C. Our model for organic synthesis on meteorite parent bodies is based on these observations, with constraints that were likely to prevail during aqueous alteration.

Although humorous, Miller and Bada's use of cocaine and glycine equilibrium is misleading, owing to hand-picked thermodynamic data and compositional constraints. Their calculated abundance of cocaine is a strong function of both $\Delta G_{\text{react}}^\circ$ and their assumed concentration of glycine. Although their methods are not referenced, it is clear that the large uncertainty in their estimated ΔG_f° for aqueous cocaine propagates directly into $\Delta G_{\text{react}}^\circ$, and causes K_{eq} for the glycine to cocaine reaction to lie between 10^{40} and 10^{54} . The law of mass action for this reaction yields

$$a_{\text{cocaine}} = K_{\text{eq}} \frac{(a_{\text{glycine}})^{10} (f_{\text{H}_2})^9}{(f_{\text{CO}_2})^3 (f_{\text{NH}_3})^9 (a_{\text{H}_2\text{O}})^{10}}$$

and it can be seen that calculated cocaine activities (a) will also be subject to 14 orders of magnitude of uncertainty. This uncertainty can be masked, and results can be calculated that are consistent with abundant cocaine, by a careful choice of the activity of glycine, given its power of ten.

To arrive at their activity of aqueous glycine, Miller and Bada need a bulk glycine concentration of 10 p.p.m., and a water to rock ratio during alteration of 0.1. Their parent-body glycine concentration exceeds by 33% the highest concentration of glycine in Murchison extracts¹². It is important to remember that amino acid concentrations reported for meteorites are actually those of acid-hydrolysed, hot-water (100 °C) extracts, in which hydrolysis increases the free amino acid concentration by as much as 800% (ref. 12). For the water to rock ratio, they assume that the Murchison parent body contained 10% by volume interstitial water, and cite DuFresne and Anders⁶ (whose paper was published seven years before the meteorite fell). In fact, DuFresne and Anders state that the Orgueil meteorite contains 10% water, but this is the amount of H_2O structurally bound in hydrated silicates, which must be a small fraction of the water present during alteration. Oxygen isotopes indicate minimum water to rock ratios of 0.5 and 0.8 for Murchison and Orgueil, respectively¹³, but these values require the assumption that the minerals analysed formed simultaneously in the same aqueous environment. This appears unwarranted given present knowledge of the paragenesis of meteorite alteration¹⁴. It is more likely that water to rock ratios during meteorite alteration were much like those in terrestrial hydrothermal systems. If so, the concentrations of many organic compounds in meteorites may have accumulated slowly

from dilute solutions in the same way that massive ore deposits form from vast quantities of solution carrying a few p.p.m. metal.

By overestimating the amount of glycine, and underestimating the water to rock ratio, Miller and Bada obtain a glycine concentration ($10^{-2.4}$) consistent with metastable equilibrium only if cocaine were found in meteorites. In our calculations, we take account of all the aqueous organic compounds for which thermodynamic data can be calculated¹⁵, place mass-balance constraints on each element, and minimize the Gibbs function for the chemical system. We calculate an aqueous glycine activity of $10^{-6.8}$ during metastable aqueous alteration of pyrene at 100 °C, 100 bars, $f_{\text{NH}_3} = 10^{-3}$, and f_{H_2} set by the nickel–bunsenite mineral assemblage (f_{CO_2} is set to $10^{1.4}$ by the phase rule). The corresponding activity of aqueous cocaine is $10^{-38.7}$, consistent with the absence of cocaine in carbonaceous chondrites.

EVERETT L. SHOCK

MITCHELL D. SCHULTE

Department of Earth and
Planetary Sciences,
Washington University,
St Louis, Missouri 63130, USA

- Shock, E. L. & Schulte, M. D. *Nature* **343**, 728 (1990).
- Wright, I. P. & Gilmour, I. *Nature* **345**, 110–111 (1990).
- Henze, H. R. & Blair, C. M. *J. Am. chem. Soc.* **53**, 3042–3046 (1931).
- Miller, S. L. & Smith-Magowan, D. *J. phys. Chem. Ref. Data* **19**, 1049–1073 (1990).
- Stull, D. R., Westrum, E. F. Jr & Sinke, G. C. *The Chemical Thermodynamics of Organic Compounds* (Wiley, New York, 1969).
- DuFresne, E. R. & Anders, E. *Geochim. cosmochim. Acta* **26**, 1085–1114 (1962).
- The Merck Index* 2420 (Merck, Rahway, New Jersey, 1983).
- Zolensky, M. E., Bourcier, E. L. & Gooding, J. L. *Icarus* **78**, 411–425 (1989).
- Shock, E. L. *Geology* **17**, 572–573 (1989).
- Shock, E. L. *Geochim. cosmochim. Acta* **54**, 1185 (1990).
- Miller, S. L. & Bada, J. L. *Nature* **334**, 609–611 (1988).
- Shock, E. L. & Schulte, M. D. *Geochim. cosmochim. Acta* **54**, 3159–3173 (1990).
- Clayton, R. N. & Mayeda, T. K. *Earth planet. Sci. Lett.* **67**, 151–161 (1984).
- Tomeoka, K. & Buseck, P. R. *Geochim. cosmochim. Acta* **49**, 2149–2163 (1985).
- Shock, E. L. & Helgeson, H. C. *Geochim. cosmochim. Acta* **54**, 915–945 (1990).

Scientific Correspondence

SCIENTIFIC Correspondence is a relative informal section of *Nature* in which matters of general scientific interest, not necessarily those arising from papers appearing in *Nature*, are published. Because there is space to print only a small proportion of the letters received, priority is usually given according to general interest and topicality, to contributions of fewer than 500 words and 5 citations, and to contributions using simple language.

If new results are being described, priority is generally given to communications that do not describe work in which the author is involved. Authors of contributions of this type should explain in a covering letter why theirs has a particular claim on *Nature's* space. Contributions may be sent to referees and, in the case of matters arising from material published in *Nature*, are sent to the author of that article for comment. A more detailed guide to authors is available from Washington or London. □

On space conservation

William J. Lennarz

Biochemistry. By Donald Voet and Judith G. Voet. Wiley: 1990. Pp:1,223. £23.50, \$39.95.

FOR someone who did his biochemistry 'teething' in the 1950s on classic texts such as that by J. Fruton and S. Simmons (*General Biochemistry*, Wiley, 1953), the modern biochemistry textbook is a thing of wonder. Having gone from Fruton and Simmons to watching at close hand A. Lehninger writing the first of a new generation of textbooks (*Biochemistry*, Worth, 1970), and finally to the stage where I see undergraduate, graduate and medical students poring over tomes such as *Molecular Biology of the Cell* by Bruce Alberts *et al.* (Garland, 1989) and *Molecular Cell Biology* by James Darnell *et al.* (Scientific American Books, 1990), I am astounded by the effective manner in which contemporary textbooks convey the vast collection of facts that comprise the current state of biochemical knowledge. This, of course, has been accomplished by the application of the old Chinese adage about one picture being worth a thousand words.

Biochemistry by Donald and Judith G. Voet, with illustrations by Irving Geis and John and Bette Woolsey, follows in the current tradition. It is lavishly illustrated — one has to search hard to find a page in this 1,223-page book that does not contain an illustration — and most of them are multi-coloured. The vast majority of these illustrations are extremely useful. Occasionally, however, an illustration is of only marginal value in the context of the text. For example, it is not clear that a detailed diagram of the substrate (lactose) and the products (galactose and glucose) is necessary in the section dealing with β -galactosidase and enzyme induction in the chapter on transcription, where it might seem merely to be a distraction. In a similar vein, filling two-thirds of a page with structures of tunicamycin, and of UDP-N-acetylglucosamine and dolichol phosphate, to illustrate the highly speculative hypothesis that this antibiotic is structurally similar to an 'adduct' of the latter two compounds is of dubious worth. I would offer the thought that perhaps textbooks of this ilk have gone too far, and that some careful editing of illustrations would if nothing else at least restrain the rate of their growth in size and weight.

A major issue with this and every other comprehensive textbook is the matter of organization. Clearly, how to deal with such a vast collection of facts is not trivial. The text is organized into five parts: Introduction and Background, Biomolecules, Mechanism of Enzyme Action, Metabolism and the Expression and Transmission of Genetic Information. As many other textbooks, the approach is to present a dull compendium of

information on the structure of the biomolecules, briefly touching on their function, but holding off discussion of their synthesis and related issues until a later chapter. Arguably, it would have been better to limit the initial coverage of biomolecules to nucleic acids and proteins, then deal with enzyme mechanisms, and then cover each class of biomolecule and cell-biological context. In this way it might have been possible to avoid unnecessary repetition, thereby saving space. A slightly better organization might also have avoided the need to hop around to obtain a unified picture. For example, as it stands glycoproteins are dealt with in chapter 10 under the sugars and polysaccharides, in chapter 11 under membrane assembly and targeting, and in chapter 21 in

a section dealing with biosynthesis of oligosaccharides and glycoproteins. It seems to me that this subject could have been covered with more economy of text and illustration at one point under the arrangement outlined above, although this would no doubt have introduced other problems. In any case it seems that the time has come when authors of such comprehensive texts must make more efforts to conserve space without overall loss of content.

In summary, this is a solid, well-written text that is quite up to date and is comparable in overall quality and in coverage to new or new editions of other comprehensive texts such as Alberts *et al.*, Darnell *et al.* and Matthews and van Holde (for a review see *Nature* 345, 30; 1990). Only time will tell whether or not *Biochemistry* will successfully compete with already established textbooks. □

William J. Lennarz is in the Department of Biochemistry and Cell Biology at State University of New York at Stony Brook, New York 11794-5215, USA.

Forensic anthropology

M. A. Green

Witness from the Grave: The Stories Bones Tell. By Christopher Joyce and Eric Stover. Little, Brown/Bloomsbury: 1991. Pp.333. \$19.95, £16.95.

FORENSIC anthropology — the scientific study of human remains — is in the United States a discipline in its own right, whereas in the United Kingdom and France it remains largely in the domain of the forensic pathologist. *Witness from the Grave* starts with a general account of the uses of forensic anthropology, with the remaining two-thirds of the book specifically detailing the identification of the remains of Josef Mengele of Auschwitz and the exhumation of those who disappeared in Argentina during the seven years up to 1983. The discussion of these two case histories are linked by the biography of Clyde Snow, whose unique contribution to forensic anthropology is internationally acknowledged.

Numerous other interesting cases are described ranging from aviation accident pathology to the horrific tale of the Des Plaines Illinois murderer who sexually molested and killed over 30 victims. The book is enlivened with light touches of humour. The tale of the wax-

work which in fact turned out to be the embalmed body of an early twentieth-century outlaw is hilarious. It is also interesting to learn that even in his declining years, Josef Mengele still carried a packet of condoms with him. Hope springs eternal!

The account of the emergence of forensic anthropology as a discipline and its value in providing evidence in court is excellent. *Witness from the Grave* lucidly explains the techniques and their limitations. The explanations are excellent and the illustrations accompanying them superb, but there is too much repetition. For example, accounts of facial reconstruction and of superimposition appear in both the early chapters and in the account of the Mengele case. ►

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Josef Mengele — detailed examination of his remains was carried out to make a positive identification.

I enjoyed reading the book, although I was irritated by the breathless, adulatory style and by the use of English so American that the old joke about 'two cultures divided by a common language' surfaces yet again. I also have reservations about the scientific value of the book for the reasons given below, but I must emphasize that the scientific data presented are both accurate and comprehensive.

Snow's experiences in Argentina are worthy of record, and it is right that the memories of this period of oppression should be preserved. Although *Witness from the Grave* has been widely researched, the references are of limited value. Many are from press-cuttings libraries, or from material that is unpublished or not 'readily available', or from personal communications. There are, unfortunately few citations of

major texts and papers.

I am unable to recommend this book to the serious scientist or academic department. The information is better obtained from the works of authors such as W. M. Krogman and K. J. Reichs, or from the mainstream journals. *Witness* will, however, be of use to lay readers with an interest in criminalistics, although they might jib at the price. Some lawyers or police officers may also find it interesting.

I hope the book does well, but unfortunately it falls between two stools; it is too general for the specialist and too wordy and expensive to have more than a limited public appeal. □

M. A. Green is at the Medico-legal Centre, Watery Street, Sheffield S3 7ES, UK.

A symbiotic relationship

Demosthenes Kazanas

Cosmic Rays and Particle Physics. By Thomas K. Gaisser. Cambridge University Press: 1991. Pp. 279. Hbk £35, \$65; pbk £15, \$24.95.

THERE can hardly be anyone in physical sciences who has not become aware of the interplay between the two diverse disciplines of particle physics and cosmology that has stormed through both scientific and popular editions in the decade past. Although the impact of the developments in particle physics on cosmology (and the other way around to a lesser degree) has no doubt been great, this interplay has been mainly confined to the purely theoretical domain (with the exception of the limits on the number of neutrino flavours set by Big Bang nucleosynthesis and confirmed by LEP at CERN). But the hard facts have been less forthcoming than one would have guessed from the flurry of theoretical model-dependent pronouncements. Although this is the case for the interplay between cosmology and particle

physics, a similar symbiotic relation between the disciplines of particle and cosmic-ray physics has been growing quietly with a lot more relevance to experimental facts, and it is the subject of *Cosmic Rays and Particle Physics* by Thomas Gaisser.

Cosmic rays have been the subject of intense scientific study during the past several decades, and it is worth remembering that several of the discoveries in particle physics were made using the cosmic-ray beam. But the advent of laboratory accelerators has led progressively to the decline in significance of the cosmic-ray beam as a means for studying high-energy physics. The emphasis in cosmic-ray studies subsequently shifted to more astrophysically oriented issues, such as the spectrum, composition, and acceleration of cosmic rays, whose resolution, incidentally, is likely to remain elusive for the foreseeable future. So at the end of the last decade, the field of cosmic-ray physics appeared mature enough to preclude the surprises and quick advances reserved for subjects of a more youthful age.

At this point the latest developments in particle physics have forced upon us several facts that have had an impact on the study of cosmic-ray physics. For example, the discovery of new quark flavours with very short life-times has suggested the presence of a prompt muon component in the atmospheric cosmic-ray cascades (that is, muons emitted from meson decay before the parent meson had suffered any collisions in the atmosphere). More importantly, the search for proton decay, motivated by theoretical considerations, has prompted the construction of a number of large underground detectors for which the atmospheric muon and neutrino-induced events produced by the cosmic radiation beam constitute the background rather than the signal. Cosmic-ray physicists have thus been forced to refine their calculations to determine this background as best as possible. To this goal they have received help from accelerator data, which can now provide information about cross sections up to ~ 1 TeV at the centre of

mass, thus further reinforcing the symbiotic relation between high-energy and cosmic-ray physics and providing an unexpected boost to the latter. It is precisely the analysis of the physics of the cosmic-ray-induced high-energy cascades and the details associated with the detection of the various components (for example, electrons, muons, neutrinos) both on and under ground that constitutes the main thrust of this book. In this respect, the material covered in the book does not contain as much novelty of notions and ideas as one might expect having been exposed to the interplay between particle physics and cosmology. But it contains a lot of material and facts that are relevant to real-life detectors and observations as they have been shaped by the latest advances in high-energy physics and astrophysical discoveries. Discussions of some more speculative, exotic stuff, such as neutrino oscillations, are included, but then only within the context of their impact on observations with and on limits set by real detectors.

The apparent absence of any significant proton decay and the nearly simultaneous discovery of point sources of very high (10^{12} eV) and ultra-high (10^{15} eV) radiation associated with compact accreting objects in our Galaxy, endowed the proton-decay detectors with the alternative function of high-energy astrophysics telescopes. This additional function did in fact cause quite a stir with the tentative report of muon-rich showers (that is, richer than expected from photon-initiated cascades) in these detectors from the direction of one of these astrophysical ultra-high-energy sources, hinting at the possibility of new physics at TeV energies. The major theoretical and experimental issues associated with these sources are also reviewed in the book. In addition, one can find a succinct but lucid exposition of a variety of other topics related to cosmic-ray research, such as shock acceleration, propagation, the role of antiprotons as probes of cosmic-ray sources and propagation and so on.

Cosmic Rays and Particle Physics offers quite a thorough overview of the physics of cosmic-ray cascades in the atmosphere and underground as they have been shaped by the needs and developments of high-energy physics in the last decade. The treatment of all the topics is concise and based on or supplemented by sufficient physical argumentation to allow a much clearer understanding of the accompanying equations. The presentation is excellent and very useful to graduate students (to whom the book is mainly addressed) and also for researchers wishing to familiarize themselves with cosmic-ray cascades and cosmic rays in general, and there are a number of reviews and references provided that should suffice as a first step for a more thorough study. □

Demosthenes Kazanas is at the NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA.

New in paperback

■ *James Joule: A Biography* by Donald S. L. Cardwell published by Manchester University Press/St Martin's Press at £12.95, \$24.95.

■ *Wonderful Life: The Burgess Shale and the Nature of History* by Stephen Gould published by Penguin/Norton at £6.99, \$10.95.

■ *The Deluge and the Ark: A Journey into Primate Worlds* by Dale Peterson published by Avon at \$11.95.

■ *Genetically Engineered Organisms: Benefits and Risks* by J. R. S. Fincham and J. R. Ravetz published by the Open University Press at £8.99.

■ *Liquid Crystals: Nature's Delicate Phase of Matter* by Peter J. Collings published by Princeton University Press/Adam Hilger at \$10.95, £9.95.

The Great Attractor: do galaxies trace the large-scale mass distribution?

Alan Dressler

Large-scale inhomogeneities have been discovered in the distribution of galaxies and attempts have been made to infer, from their motions, the distribution of the additional, invisible matter that dominates the dynamical evolution of the Universe. Like the galaxy distribution, the smoothed distribution of mass density shows high-contrast features stretching over vast regions. The local volume of the Universe seems to be dominated by a single region of above-average mass density: the Great Attractor.

TRADITIONALLY, 'real' physicists have cast sceptical eyes toward observational cosmology. With little more than bits of glass and wire to capture the echoes of remote catastrophes, the argument goes, what chance have we to sort out the story of creation itself?

It is one thing to collect data. After all, the Universe is transparent, and a consequence of the finite speed of light is that the record of all history is an open book. We can see galaxies, measure their motions, assay their chemistry, and can, in principle at least, do this for epochs reaching back billions of years.

Interpretation may be another matter. Edwin Hubble, an astronomer of legendary accomplishments, was even reluctant to interpret galaxy redshifts as evidence for an expanding universe, a picture embraced enthusiastically by general relativists of his day. Wary of the leap from observation to theory, he cautioned "Not until the empirical sources are exhausted need we pass on to the dreamy realms of speculation". Indeed.

That cosmology was bogged down in the 1970s was, therefore, in keeping with the general hopelessness of the task. Observers had been unable to tie down the expansion rate (and thus the age) of the Universe to better than a factor of 2, and a monumental effort to find out if the Universe would recycle or expand forever had stalled¹. Theorists seemed to have no clue as to what had initiated the Universe's dynamic expansion, or how the process could have produced fluctuations in the density of matter large enough to lead to galaxies of stars.

It was, then, rather a surprise when the 1980s proved to be a renaissance for cosmology, with observational astronomers, theoretical astrophysicists and even particle physicists rebounding with new data and fresh ideas. Driving the change were two factors: the discovery that the distribution of galaxies in space is complex and highly structured—more like sculpture than spatter—and the development of the idea that the Big Bang began with a furious episode of inflationary growth. The thought that such structure might be the imprint of early-Universe physics, and indeed might be the most direct path to understanding physics in this remote realm, pumped new vigour into the discipline.

Cosmology is now one of the liveliest fields in astrophysics. Research centres on (1) characterizing the distribution of galaxies; (2) characterizing the distribution of the mass, which is not necessarily the same as (1); and (3) evolving model universes from different physical bases and comparing them with observations. In this article I review the impact that the discovery of the Great Attractor, an enormous concentration of invisible matter, has had on these areas.

By the late 1970s, astronomers were making the first detailed maps of the distribution of galaxies. They added the critical third dimension by acquiring large numbers of galaxy redshifts and applying Hubble's discovery that recession velocity is approximately proportional to distance. Slowly but surely, the

half-century-old dogma that the distribution is smooth on large scales, interrupted only by occasional concentrations of the galaxies into clusters, began to give way to persuasive evidence for large, nearly empty regions with surprisingly thin sheets and filaments of galaxies wrapped around them. The publication of the first of the redshift survey 'slices' by the Harvard-Smithsonian Center for Astrophysics which mapped the positions of 1,200 galaxies in a thin, fan-shaped cut through the sky, provided the most dramatic evidence for an intricate, high-contrast structure.

This discovery raised important questions on how the Universe, whose early homogeneity could be inferred from the uniformity in the cosmic microwave background could have evolved in 10–20 billion (thousand million) years to its present strikingly inhomogeneous appearance. Current models in the field of high-energy particle physics suggest that the seeds of structure may have been planted at the earliest moments of the Big Bang, and hence the study of large-scale structure may offer a critical insight into the process of creation itself.

However, an assumption essential to this endeavour is that galaxies are accurate indicators of the large-scale mass distribution, believed by most researchers to be primarily composed of non-radiating and possibly non-baryonic matter. Fortunately, we can chart the distribution of all matter, whether visible or not, by studying not just positions but motions of galaxies, because the Hubble expansion flow will itself be perturbed by a lumpy distribution of matter which has an undulating gravity field. As a result of a survey of the properties of elliptical galaxies, seven of us were able to make an omni-directional map of the mass distribution over the largest volume yet attempted. The large coherent flow of galaxies (including our own Milky Way) that we found suggests the presence of an extensive overdensity of matter, which we call the Great Attractor, and reinforces the conclusion that the distribution of all matter in the Universe is inhomogeneous on very large scales.

The discovery of large-scale structure

We interpret the observed redshifts of galaxies as Doppler shifts caused by their motions in an expanding universe; specifically, the redshift of a galaxy is roughly proportional to its distance from the observer. This provides a relatively easy way to go from the two-dimensional projection of galaxy positions in the night sky to true space dimensions. In the 1980s, advances in telescope detector technology and the tenacity of a group of Harvard-Smithsonian astronomers expanded the number of galaxies with known three-dimensional space positions to several thousand². One special survey concentrated on a narrow strip of sky, providing the densest sampling of space yet attempted. The results, reported by de Lapparent, Geller and Huchra³, showed a tapestry-like pattern, with long chains of galaxies in thin contours surrounding large empty regions nearly devoid of galaxies. Further measurements of redshifts for

galaxies in this region of the sky have shown one of the chains to extend well above and below the original survey slice, forming a 'Great Wall' of galaxies hundreds of millions of light years in length, but only tens of millions of light years thick⁴.

Can we explain the occurrence of such complex structure? A widely accepted picture is that large-scale clumping of the universe is the result of gravitational instability. The extreme smoothness of the 2.7 K cosmic microwave background, which emanates from the epoch of decoupling of matter and radiation 100,000 years after the Big Bang, suggests that the Universe was extremely uniform in its youth⁵. Nevertheless, small primordial density variations would have grown in cosmic time as mutual gravitation drew more and more matter into overdense regions at the expense of underdense regions.

This 'gravity amplifier' would not have been strong enough to turn the quantum fluctuations expected in the Big Bang plasma into today's galaxies and clusters, given the limited time (10–20 billion years) of the expansion phase. However, a very early and sudden amplification of these fluctuations, as predicted by the inflationary universe model developed by Alan Guth⁶, could have done the job. Guth investigated the effect of a first-order phase transition in the primordial soup, in which an enormous release of energy accompanied symmetry-breaking of the Higgs field that represented unification of the strong, weak and electromagnetic forces at high energy. Such energy could have propelled the Universe into rapid expansion, inflating each dimension exponentially some 40 orders of magnitude in only $\sim 10^{-32}$ s. Set into motion with a near-perfect balance of gravitational and kinetic energy, the Universe would have been left coasting, as we find it today.

In addition to identifying a cause for the expansion, the inflationary universe model could solve other riddles. It could explain how the energy content in the cosmic microwave background can be so uniform over the sky, implying that the entire Universe was once causally connected, whereas today the Universe appears far too big for a signal ever to have propagated from one side to the other. Most importantly for the present discussion, inflation offers a mechanism which can increase the size of adiabatic fluctuations enough to seed large-scale structure. (Adiabatic fluctuations are those for which the ratio of matter density to radiation density is constant over space across regions where the total mass and energy densities vary significantly. They are to be contrasted with isothermal fluctuations, which have a homogeneous distribution of radiation regardless of fluctuations in the matter density.) The inflation model even predicts a precise spectrum of fluctuations which serves as a starting place for models of growth based on gravitational instability: this is the Harrison–Zel'dovich form, known as the scale-invariant spectrum because the amplitude of a characteristic fluctuation is the same for each mass scale as it first comes within the event horizon.

With this input spectrum and the simple physics of gravity, a computer can create model universes, employing a numerical code that reduces the insoluble N -body problem to a series of tractable approximations. Such models were remarkably successful at producing structures resembling those found in the real Universe^{7,8}. Galaxies could be identified in the simulations, and they appeared to clump into small groups and rich clusters in a way more-or-less faithful to the observations. Even some of the properties of galaxies themselves seemed to fall out naturally. But for structure on the largest scales, superclusters (clusters of clusters) and voids, the model universes seemed rather too smooth, the galaxies too uniformly arranged. The thin shells and filaments outlined by galaxies, and the huge low-density regions they surround, were at best a stretch for the models.

As N -body simulations are most accurate when modelling density contrasts of less than one order of magnitude, this failure to match structure on the largest scales was disturbing. Furthermore, on such large scales gravity should be the only important

mover—the energy released by star formation in young galaxies or the effects of gaseous dissipation should be negligible. The disagreement seemed real, but with only qualitative comparisons between models and data, the interpretation remained inconclusive. Proponents hailed the models as remarkable for explaining so much of the structure, sceptics focused on the apparent failure to describe structure on the largest scales⁹.

In truth, this comparison of models and data was not an entirely valid test. The N -body models actually predict the distribution of dark matter, not galaxies. There is now persuasive, if not compelling, evidence that the dominant form of matter in the Universe is not the ordinary proton–neutron–electron matter from which we and the Earth, Sun and stars are fashioned, but a type that does not form high-density configurations. The rapid motions of galaxies in groups and clusters convince us that there is considerably more matter between galaxies than can be attributed to the baryonic material in the galaxies themselves, and the remarkably successful model of Big Bang nucleosynthesis¹⁰ indicates that this unseen extra matter, probably 10–100 times as much as we see in the galaxies, is almost certainly non-baryonic. In fact, the inflationary model almost demands the presence of dark, exotic matter, because it predicts a universe with a density high enough for gravitational energy to balance the kinetic energy of expansion, a density far greater than that of the visible baryonic matter.

What, then, is it? The most popular version invokes an invisible 'gas' of particles that interact weakly with ordinary baryonic matter because they are immune to the strongest force in nature, the electromagnetic force. Most N -body simulations have been based on the model of cold dark matter (CDM), so named because it postulates massive particles which move slowly compared with the speed of light¹¹. Another possibility often discussed is that one of the three kinds of neutrinos might have a small rest mass (some tens of electron volts) and thus be a candidate for the invisible mass. Because such neutrinos would move at relativistic speeds, this model has been given the name hot dark matter (HDM).

Though unable to produce anything really solid, such a sea of dark matter particles would nevertheless interact through gravity. Gentle undulations of its high mass density would dominate the fabric of space-time. Though the whole notion might seem outlandish, it is entirely reasonable that a fraction of the energy unleashed in the Big Bang was spent in producing such particles. Still, we cannot at this point demonstrate the existence of such exotic particles by direct means (for example, by trapping one as it passes unblinking through the Earth, or producing one in a particle accelerator). This, therefore, is one of the great attractions of comparing computer models with

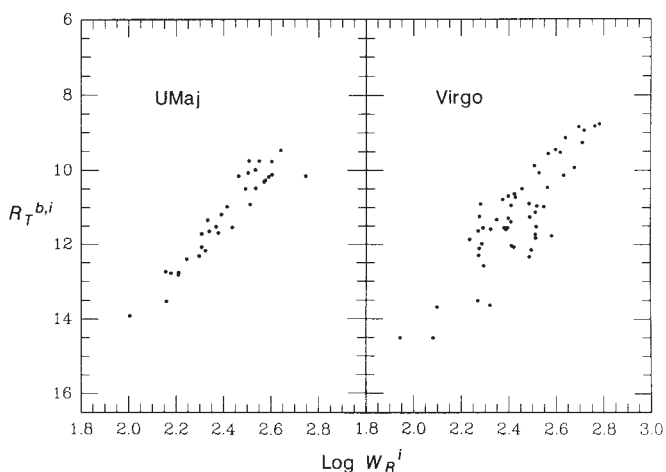


FIG. 1 The Tully–Fisher relation of rotation speed versus log luminosity for galaxies in the Virgo and Ursa Major clusters (adapted from ref. 5).

observed large-scale structure—a test for the existence of matter that gives off no light, but whose gravitational signature is as distinctive as back-slanted handwriting.

The distribution of dark matter

It has been conventional to assume that galaxies trace this invisible substructure, accurately following its peaks and valleys. In this picture, the gravity field is, and has always been, dominated by dark matter, so where a high-density dimple formed, baryons flowed in like water to a pothole and produced a galaxy on the spot. A crease in the dark matter might develop into a chain of galaxies, a giant sinkhole might encourage the formation of a cluster or supercluster. However, this is just a model, and the supposition that galaxies are accurate tracers of the mass is, at present, no more than that. It would be far better to map the distribution of dark matter directly and then compare with the predictions of N -body models. As a dividend, comparing the galaxy and dark matter distributions will test whether galaxies are fair tracers, which will itself be revealing of the process of galaxy formation.

Fortunately, nature has provided a way to map the mass distribution directly: although the major component of galaxy motion is due to the expansion of the Universe, a nearby overdensity of dark matter will pull galaxies, imparting additional velocities. Thus, galaxies will stream towards overdense regions and away from underdense regions. A map of the motions of galaxies, then, with the regular pattern of expansion velocities removed, will reveal where the dark matter is, and how great are the swells and troughs of its density. The higher the waves in the sea of dark matter, the more vigorously will galaxies be cast to and fro. Averaged over large scales, these motions are still small enough for the perturbations to be modelled with linear perturbation theory, which simplifies the recovery of the underlying mass distribution.

Observations of the expansion field had been used to argue that such 'peculiar velocities'—perturbations from a smooth Hubble flow—must be small, because many fewer galaxies are observed to be approaching our Galaxy than would be the case if the flow were 'hot' (noisy). Such reasoning missed the possibility that the flow could be locally 'cold', but like a flotilla in tight formation, galaxies could be cruising downstream together. As it turns out, this is exactly the case.

Rubin *et al.*¹² made the first concerted effort to measure departures from uniformity in the Hubble flow by measuring the motions of hundreds of galaxies in various directions, galaxies chosen to be roughly equidistant from the Milky Way. Rubin *et al.* interpreted an asymmetry they measured in the velocity field as due to a motion of our own Galaxy of $\sim 500 \text{ km s}^{-1}$, a speed which was unheard of according to conventional wisdom. All doubt that the Milky Way was, in fact, in rapid motion (although not in the direction that Rubin *et al.* had found) was removed a year later in 1976, when an unmistakable anisotropy was detected in the temperature of the cosmic microwave background^{13,14}. The background radiation streams to us directly from a time only some hundred thousand years after the Big Bang and serves as a kind of cosmic backdrop against which motions of local galaxies can be referenced to the Universe at large. This distortion in the temperature of the microwave background, interpreted as the Doppler shift caused by a peculiar motion of our Galaxy^{15,16} at 600 km s^{-1} , was the first indisputable evidence for large departures from a smooth Hubble flow. Its detection raised new questions. Over what range are other galaxies sharing this motion? Towards what are they streaming? These questions were the first steps towards the Great Attractor.

Measuring galaxy distances and peculiar motions

If we could only drop into neighbouring galaxies and measure the anisotropy of the cosmic microwave background in their night skies, our task would be simple. Unfortunately, even vast

increases in NASA's budget would not ferry us to other galaxies at any time soon, so we must measure the motions of galaxies from our restricted vantage point on Earth here in the Milky Way. Actually, it is quite simple to measure galaxy motions, at least the component towards or away from us. One collects the light of a galaxy and computes the Doppler shift by comparing the wavelengths of absorption lines in the integrated light of stars to their laboratory values. Alas, most of this measured velocity is due to the *expansion* of the Universe; the peculiar motion is a small part that can be found only after subtracting the cosmic contribution. This step requires a knowledge of the distance to the galaxy, because velocities in a Hubble flow unperturbed by a lumpy mass distribution are a linear function of distance.

Methods used at present to estimate distances to galaxies are barely adequate to the task. We cannot employ the approximation that distance is given by redshift, as this assumes that the peculiar motion, the quantity we are after, is zero. A typical random error of 20% means that for a galaxy whose distance would correspond to a recessional velocity of $5,000 \text{ km s}^{-1}$ in an unperturbed Hubble flow, the subtraction of cosmic contribution will result in an uncertainty in the peculiar motion of $\sim 1,000 \text{ km s}^{-1}$. With present techniques, then, we must measure the motions of many galaxies in a particular region of space and combine the results to obtain the needed accuracy of a few hundred kilometres per second.

These methods amount to trying to read the wattage on a distant light bulb, then calculating how far away the bulb is by measuring how dim the light appears. Of course, for galaxies the 'wattage' is actually the total luminosity of each galaxy's stars. Tully and Fisher¹⁷ showed that for spiral galaxies the total luminosity is well correlated with galaxy rotation speed. Optical or radio telescopes operated in a spectroscopic mode can measure galaxy rotation as differential Doppler shifts across the face of the galaxy.

The monotonic increase in luminosity as a function of rotation speed has been tested and calibrated using galaxies in clusters (Fig. 1). With all objects at the same distance from us, relative *apparent* brightness indicates true relative luminosity. Tests on some dozens of clusters show that rotation speed is correlated with galaxy luminosity with an r.m.s. scatter of $\sim 40\%$. By the inverse square law for light intensity, this translates into an error of $\sim 20\%$ in the distance of each galaxy. In this way the rotation speed, measurement of which is independent of a galaxy's distance, is used to predict a distance-dependent quantity, the apparent brightness.

We do not yet have a precise physical explanation for the Tully-Fisher relation, or for the similar correlation used for elliptical galaxies in which stars display more random, virial motions instead of organized circular motion. However, the basic idea is simple: stars in a galaxy move in response to gravity, which is a function of mass. As most of the mass in the visible part of a galaxy is in stars, the mass measured by kinematic means is well correlated with the number of stars and their total light output. Nature has graciously cooperated by suppressing the effect of variations in galaxy size, but has limited the precision of the correlation by making it difficult to assess the contributions of smaller determinant factors such as variations in the stellar population mix.

Maps of peculiar velocities

In 1982 the Tully-Fisher relation was applied to a sample of spiral galaxies in a fundamental study of the peculiar motions of galaxies in the Local Supercluster—the region of high galaxy density, roughly 100 million light years across, in which we live. Aaronson *et al.*¹⁸ measured distances to several hundred spiral galaxies and subtracted the expansion component from the observed recession velocity, leaving a field of peculiar motions. Their pioneering study showed a clear pattern of 'infall' of galaxies toward the region of highest galaxy density. Actually,

the Hubble expansion velocity is larger than the infall, so it is more correct to say that expansion of the Local Supercluster is retarded relative to the average expansion rate for the Universe. Outside the immediate vicinity of the Virgo and Ursa Major clusters, which are connected by a ridge and define the centre of the Local Supercluster, galaxies are still receding from each other. The study confirmed that the higher local density of galaxies did indeed outline a region of higher mass density, and that the effect of gravity over the denser region could be seen. These first measurements of large-scale gravity confirmed that there is at least 10 times as much dark matter as matter that can be seen in the visible galaxies.

But this triumph did little to identify the source of the Milky Way's motion of 600 km s^{-1} with respect to the cosmic microwave background. The component that could be attributed to the relative contraction of the Local Supercluster is only $\sim 250 \text{ km s}^{-1}$, and is in a direction quite different from that implied by the microwave dipole anisotropy. The measurements of Aaronson *et al.* suggested that the major component was yet to be found and that the Local Supercluster was moving *en masse* towards some more distant point. Was this the result of many small motions combined, or was it the influence of a Great Attractor that dominated the local velocity field?

In the early 1980s, a group of astronomers came together to work on the nature and distribution of elliptical galaxies. Sandra Faber (University of California, Santa Cruz), Roberto Terlevich (Royal Greenwich Observatory), Roger Davies (Oxford), David Burstein (Arizona State), Donald Lynden-Bell (Cambridge), Gary Wegner (Dartmouth) and I were drawn to study elliptical galaxies because of their relative simplicity. Compared with spiral galaxies, ellipticals are more spherically symmetrical and are composed of a more uniform population of old stars. We believed these factors would permit a reliable method of predicting distance. The Faber-Jackson method then in use was only accurate to $\sim 35\%$ in distance, too crude for effective mapping

of the field of peculiar velocities¹⁹. This method resembles the Tully-Fisher relation for spirals except that for ellipticals, in which the stars have more disordered motions, 'velocity dispersion' is measured rather than rotation speed. (Measurements are again made spectroscopically, with the width of absorption lines indicating the characteristic stellar speed²⁰.)

Our group of seven measured the velocity dispersions of some 400 elliptical galaxies selected by uniform criteria over the entire sky²¹⁻²³. For each galaxy we also made a photometric observation, recording the total light intensity from the object and variation of its intensity with radius. We discovered that in a three-space of velocity dispersion, total brightness and average (surface) brightness, the distribution of elliptical galaxies is nearly planar^{24,25}. This allowed us to construct a new distance indicator which, like the Tully-Fisher relation, used a distance-independent parameter, velocity dispersion, to predict a linear combination of parameters that are distance-dependent. The accuracy of this method for predicting distances was about the same as the method for spirals, $\sim 20\%$.

When we applied the method to the sample of elliptical galaxies, the result was surprising. A map of the peculiar velocities exhibited a large-scale flow pattern over a region nearly five times the size of the Local Supercluster²⁶. We confirmed that galaxies within the Local Supercluster were moving with our Galaxy, and there was no obvious end to the flow, as would be expected if the source of the gravitational pull had been reached.

We first characterized this result by the term 'bulk flow', meaning that, to first order, everything in the sampled region was moving together²⁶. This alone was in conflict with *N*-body simulations based on the cold dark matter model, which predicted a much shorter distance scale over which peculiar velocities would be well correlated. Later we began to reassess the statistical significance of our result and became convinced that the data had sufficient accuracy and leverage to justify fitting terms of higher order than the simple dipole field. In particular, it seemed significant that in galaxies in the direction of the flow ahead of the Milky Way were moving more rapidly, typically at $\sim 1,000 \text{ km s}^{-1}$. This included the rich Centaurus clusters, once considered a likely source of the flow, which were now found to be racing ahead to some more distant gravitational centre. Conversely, galaxies behind us were moving in the same direction but more slowly than the Milky Way and its companions.

This gradient was easily interpreted in terms of a quadrupole contribution to the velocity field, the expected effect of the pull of a distant mass concentration. The perpendicular component of this quadrupole term had previously been discovered in the velocity field of the Local Supercluster²⁷. This evidence encouraged us to model the flow pattern expected for a gravitational source just outside the well-sampled volume. We found that a distributed spherical mass in the direction Galactic longitude 307° , Galactic latitude 9° , whose centre lay at a distance corresponding to a recessional velocity of $4,350 \text{ km s}^{-1}$ from our Galaxy, would induce a peculiar velocity field that matched the observations in surprising detail²⁸. Normalizing the model to the observed velocities implied a mass of 5×10^{16} solar masses, a huge supercluster. This became known as the Great Attractor (GA) model. The addition of a handful of free parameters greatly reduced the chi-squared value over the 'bulk flow' solution. Within the errors of the elliptical galaxy sample, further refinements in the model were unjustified.

It is curious that not until this late stage did we think to turn back to the sky to see what is visible in the direction of the Great Attractor. An all-sky galaxy map (Fig. 2) produced by Ofer Lahav (Cambridge) from three galaxy catalogues showed that the GA region, although partly obscured by dust from the plane of our own Galaxy, is the most luminous part of the sky. Apparently a substantial overdensity of galaxies accompanies the mass overdensity revealed by the measurements of peculiar

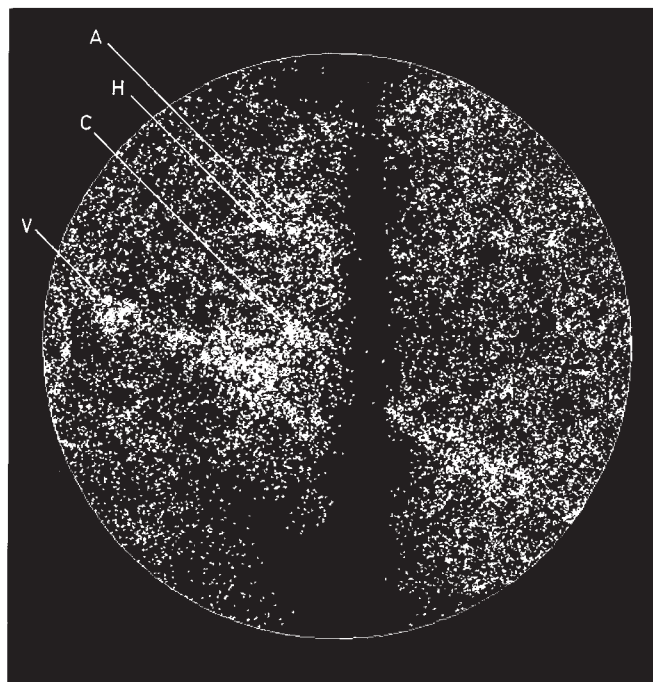
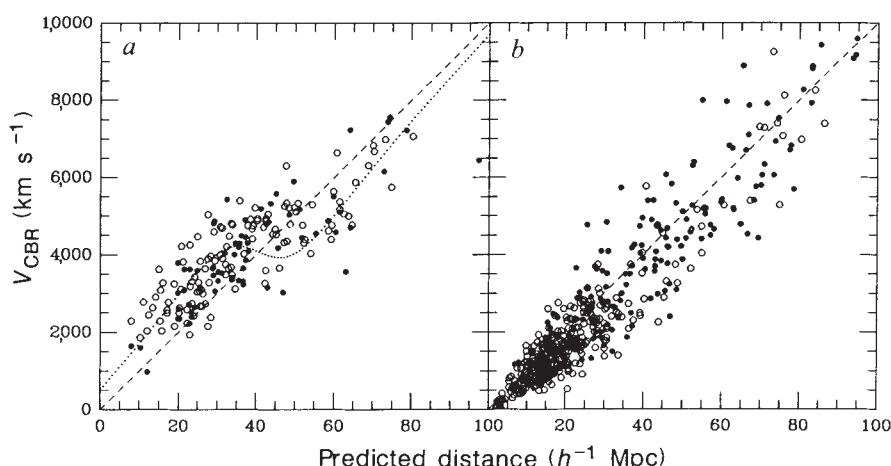


FIG. 2 Computer plot of one hemisphere in equal-area projection of the sky made by Ofer Lahav from galaxy catalogues. The picture is centred on the direction of local streaming. The dark band where no galaxies are recorded is the result of the absorption of light by dust in our Galaxy. The Virgo, Centaurus, Hydra and Antlia clusters are indicated by their initials, but the greatest concentration of light is a larger, more diffuse feature just below the Centaurus cluster. This overdensity of galaxies is associated with the Great Attractor.

FIG. 3 *a*, Hubble diagram of predicted distance versus observed velocity for the combined Dressler-Faber data sets of elliptical and spiral galaxies in the GA region. The S-curve, a GA model (and not a fit to the data) by S. M. Faber and D. Burstein, is in good agreement with the data, especially considering that most of the original data on which this prediction was based came from galaxies on only one side of the GA at distances of 20–60 Mpc from its centre. *b*, Hubble diagram for elliptical and spiral galaxies outside the GA region (over the rest of the sky). Despite identical treatment of the data, the contrast between this Hubble diagram and the S-wave seen in *a* is striking.



motions. At least in a rudimentary sense, galaxies were found to trace the hidden mass.

The GA model offered a serious challenge for the CDM theory. *N*-body simulations by Bertschinger and Juskiewicz²⁹ showed that such a massive and extensive concentration was extremely improbable, especially in the commonly assumed case where galaxies were supposed to cluster more strongly than the dark matter ('biasing'). They pointed out, however, that the GA model was not well constrained by the data then available, and that a number of less extreme mass distributions would also fit the data. Further tests of the GA model could clearly provide a strong constraint on CDM.

Many subsequent studies, then, aimed at confirming the GA model. I conducted a redshift survey of 1,300 galaxies in the GA region which confirmed that most of the galaxies seen in the Lahav map had recessional velocities of 2,000–6,000 km s⁻¹ (refs 30, 31). There is, indeed, an overdensity of galaxies relative to the average density in the local Universe. Assuming that galaxies were fair tracers of the mass distribution, this was quite consistent with the GA model.

More important, however, were new measurements of peculiar motions. Large-amplitude peculiar motions were found in the GA region by Aaronson *et al.*³² for cluster spirals, and by Lucey and Carter³³ for cluster ellipticals. Although these independent observations generally supported the GA model, they also uncovered anomalous motions which indicated, not surprisingly, that it was oversimplified.

Sandra Faber and I spent three more observing seasons at the Carnegie Institution's Las Campanas Observatory obtaining data for 134 elliptical galaxies in the GA region and 117 spiral galaxies, both in and out of clusters^{34–36}. The new data not only represented a manifold increase in the number of galaxies in the region with known peculiar velocities, but also probed further into and beyond the putative centre of the GA. This is crucial because the unambiguous signature of a region in gravitational contraction is a change in sign of peculiar velocities as the centre of mass is crossed. The new data (Fig. 3*a*) show just such an effect. Compared with the straight line that indicates uniform Hubble expansion, galaxies on the near side of 4,000 km s⁻¹ are moving too fast for their distance: they have a net positive peculiar motion. Beyond this the peculiar motions drop rapidly, crossing zero at a 'distance' of ~5,000 km s⁻¹ and then turning negative. Although the most distant points are subject to greater noise and increasing systematic errors, it would appear that galaxies on the far side of the Great Attractor are 'falling back'. A completely independent data set for spirals by Mathewson and collaborators (private communication) shows the same pattern. The full S-wave pattern, so named because the data describe a full oscillation around the Hubble line, is the expected signature for a region in which the Hubble expansion has been slowed.

The reality of the S-wave observed in the GA region can be judged by comparing with the Hubble diagram done in exactly the same way for galaxies in other directions in the sky: this distribution (Fig. 3*b*) is more like an unperturbed Hubble flow, with no evidence for any large gravitational distortion.

The picture as it now appears is summed up in Fig. 4. This is a map of the peculiar motions of galaxies in a volume of space generally aligned to the direction of local flow. The vectors showing galaxy motion are colour-coded to show that the large-scale flow pattern towards the Great Attractor is the dominant feature. The tidal field is also quite visible.

Do galaxies trace the mass?

Although the GA model is undoubtedly an oversimplification, the data do support the notion that this one large overdensity dominates the velocity field over a region of space ~400 million light years across. This is still a small volume of the Universe, much less than 1%, but it does represent a significant extension to our knowledge of large-scale structures. Although of similar scale to the Great Wall and other prominent superclusters, the Great Attractor is unique in that it is the first large-scale feature mapped in the mass distribution, the more fundamental quantity of large-scale structure.

It is important to emphasize that although it was clear from 1986 that the Great Attractor might amount to nothing more than the dark matter associated with several superclusters of galaxies in this region of the sky, this could not be concluded without prejudice until an extensive galaxy map had been completed. The value of the GA model was to describe the mass distribution *independently* of the galaxy distribution, as a step towards an unbiased assessment of whether galaxies actually trace the distribution of dark matter. In October 1990 the British press carried a report of the 'death of the Great Attractor', citing a paper by Rowan-Robinson and collaborators³⁷ in which they conclude that luminous galaxies in the region, if embedded in a dark matter distribution at the density necessary to 'close' the universe are sufficient to explain the peculiar motions of galaxies. This work, although an important step in comparing the maps of galaxies and mass, misrepresented the concept of the Great Attractor as due to dark, mysterious matter unassociated with galaxies, and thus failed to emphasize the important point that light seems to trace mass throughout the volume of space considered in their survey.

Because present techniques limit the accuracy of distance measurements to 20% per galaxy, it is difficult to determine peculiar velocities in the more distant Great Wall and judge whether the underlying mass structure is similar to the visible structure. No doubt there are many other 'great attractors' which are just beyond the reach of present techniques. At present, however, we must focus our attention on the smaller volume

within $6,000 \text{ km s}^{-1}$ to compare the distribution of light and mass.

Comparisons have been made with galaxy samples from catalogues using visible and far-infrared light, the latter from observations made with IRAS (the Infrared Astronomical Satellite). Redshifts for several thousand galaxies detected by IRAS have been used to construct galaxy maps, with the assumption that this subset is representative of *all* galaxies within the volume^{38,39}. The IRAS survey has certain advantages over optical surveys: a single instrument was used for the entire sky, and observations in the far infrared are less affected by dust absorption in our Galaxy. The results from two different IRAS samples and the optically-selected sample vary somewhat, but a basic feature of all surveys is that there are two main galaxy concentrations in the volume, one associated with the Great Attractor and another identified as the Perseus-Pisces supercluster. Estimates of the relative strength of these concentrations range from near parity to a 2:1 ratio in favour of the Great Attractor.

The issue of whether light traces mass in the volume may rest on whether the Perseus-Pisces supercluster is as rich in galaxies as the Great Attractor region. Figure 5 shows a map of the mass distribution calculated from the velocity field⁴⁰ compared with one of the IRAS samples. The disagreement is striking. The galaxy map implies that we are caught midway between two comparable overdensities, yet the peculiar velocity field shows no such bimodality—a single mass feature appears to dominate. However, another IRAS sample shows the effect less strongly, which indicates that we have not yet obtained the definitive galaxy sample. Furthermore, data for peculiar motions in the Perseus-Pisces region remain sparse, particularly in the densest concentrations in the direction of the Perseus constellation.

Recently, Willick⁴¹ has presented preliminary evidence that galaxies on the Pisces side of the supercluster exhibit only a minor infall pattern—less than half as strong as that of the Great Attractor—and that the entire Pisces region is moving in the

general direction of the Local Supercluster/Great Attractor at $\sim 400 \text{ km s}^{-1}$. If this result holds, and is supported by measurements now being processed for galaxies in the Perseus region, it will exacerbate the difference between the distributions of light and mass. Furthermore, as the pull of the Great Attractor, even combined with the Local Supercluster, should be too weak to pull Perseus-Pisces at this speed, such a result may imply that a significant part of the motion is due to a yet more distant and greater attractor^{42,43}.

Within the next five years, it will be possible to assemble a truly representative galaxy survey and to extend measurements of peculiar motions enough to test the degree to which galaxies trace the mass distribution, and the degree to which the mass distribution is dominated by large concentrations like the Great Attractor. If it turns out that galaxies are good tracers of the underlying mass distribution, we can proceed with confidence to study large-scale structure solely from the more easily obtained distribution of galaxies. But there is some early evidence that galaxies trace the mass poorly, and in a way that may not be easily predictable. Such an important claim requires much better proof, but if true, this will complicate or even block attempts to read and interpret large-scale structure of the universe solely from galaxy maps. On the other hand, discovering that galaxies do not follow precisely the distribution of dark matter would be a powerful clue to the questions of how galaxies formed and how structure was seeded in the early Universe.

Future directions

Observations of many types and from many sources continue to indicate that large sub-units, at least 500 million light years across, are a common feature of the universe. Even more radical descriptions may lie ahead: a recent report by Broadhurst *et al.*⁴⁴ suggests a surprisingly distinct preferred scale and even a hint of a periodic structure. Such measurements may have already forced the simple CDM model out of the running. Starting with the expected spectrum of fluctuations and the

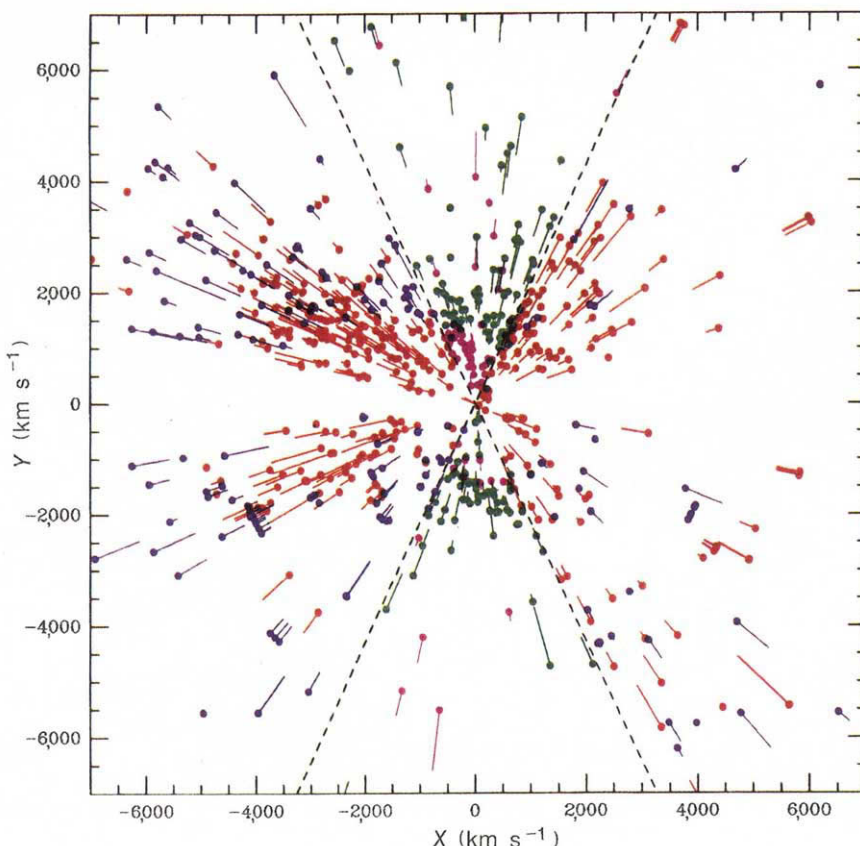
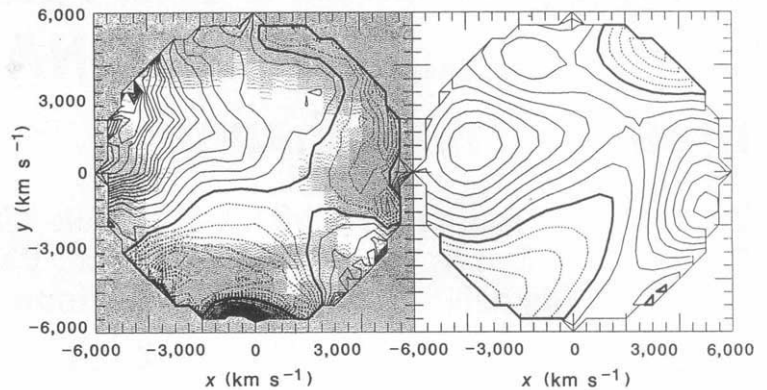


FIG. 4 Map of peculiar motions of galaxies in a volume of space generally aligned with the direction of local flow. Each vector represents a galaxy (or the median motion of a group of galaxies) with the tail showing its location in space and the arrow length showing the amplitude of its motion. Vectors coded in red show galaxies moving towards the Great Attractor centre at $\sim 5,000 \text{ km s}^{-1}$ distance directly to the left. 'Infall' can be seen from the blue vectors further to the left which point back in towards the GA centre. In the zone marked by the narrow triangles it is impossible to detect motion towards the Great Attractor because this motion would be across the line of sight and not produce a Doppler shift. Here galaxies moving towards our Galaxy are coded in green and those away in magenta. The dominance of green vectors shows the 'tides' raised by the Great Attractor, and also the contraction of the Local Supercluster, particularly obvious as the two-colour patch immediately above our Galaxy's position at the origin.

FIG. 5 The mass distribution inferred from the light distribution, as traced by galaxies detected by the IRAS satellite (right), compared with the mass distribution calculated from the measurements of peculiar velocities (left). The maps are shown in the form of density contours, with greater-than-average density indicated by solid contours and less-than-average density by dotted contours; average density is marked by the heavy contour. The shaded region of the left map indicates areas of greater uncertainty due to sparse sampling of the field of peculiar velocities. The IRAS map shows two peaks of high density, one on the left corresponding to the GA region and one on the right corresponding to the Perseus-Pisces supercluster. In contrast, the mass map derived from measurements of peculiar velocities is completely dominated by the GA; at the location of the Perseus-Pisces peak in the IRAS distribution, the mass map actually shows an underdense region. This apparent contradiction is not proven because as yet there are relatively few measurements of peculiar velocity in this region, as indicated by the shading.



amplitude implied by present small-scale clustering, it seems difficult to grow these very large structures in the time since the creation of the universe. The HDM model appears to suffer from the same dilemma: any one solution does not provide a good match for the structure on both large and small scales. Other models which rely on initial non-gaussian perturbations, such as cosmic strings or explosions, might fare better, but because the theory underpinning these models is more abstract, it is more difficult to perform a decisive comparison of models and data⁴⁵.

Of course, such confrontations between scientific theory and observations have traditionally spurred new thinking that moves a field forward, and in a sense, they should be welcomed. Nevertheless, there is an increasing sense of foreboding that there may be a serious contradiction in our observations of the evolution of structure. As discussed earlier, the cosmic microwave background, which paints a picture of the universe when it was very young, is remarkably smooth. Recent measurements of microwave temperature fluctuations⁴⁶ on scales of 0.1–1.0 degrees set upper limits for r.m.s fluctuations at a level $\Delta T/T \leq 3 \times 10^{-5}$, yet straightforward and robust calculations show, for example, that the fluctuation that grew into the Great Attractor

would have generated a $\Delta T/T \approx 10^{-5}$. Basically, photons lose this much energy as they climb out of the gravitational potential well of the embryonic Great Attractor.

The present failure to detect the fluctuations that grew into galaxies, clusters and superclusters still leaves acceptable models, but many have already bitten the dust^{47,48}. The whole notion that early fluctuations grew by gravitational instability into today's structure could well be invalidated if upper limits on temperature fluctuations are pushed a factor of five lower. Exotic models have been proposed as replacements: for example, cosmic strings, domain walls or the idea that a structure was seeded by particles which decayed *after* the decoupling epoch⁴⁹. However, because these ideas are less well developed and require 'tuned' parameters, they might be viewed as a retreat in the battle to understand the growth of structure and early-universe physics. Astronomers and physicists alike are keeping a nervous eye on the measurements of fluctuations in the temperature of the cosmic microwave background, hoping that a decade of rapid advance is not leading us up a blind alley. □

Alan Dressler is at the Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, California 91101, USA.

1. Sandage, A. & Tammann, G. A. 1984, in *Large Scale Structure of the Universe, Cosmology and Fundamental Physics* (eds Setti, G. & van Hove, L.) 127–151 (CERN, Geneva, 1984).
2. Huchra, J. P., Davis, M., Latham, D. & Tonry, J. *Astrophys. J. Suppl. Ser.* **52**, 89 (1983).
3. de Lapparent, V., Geller, M. J. & Huchra, J. P. *Astrophys. J.* **302**, L1–L5 (1986).
4. Geller, M. J. & Huchra, J. P. *Science* **246**, 897–903 (1989).
5. Wilkenson, D. T. in *Large Scale Structure of the Universe, Cosmology and Fundamental Physics* 103–169 (CERN, Geneva, 1983).
6. Guth, A. H. in *Inner Space, Outer Space* 286–301 (University of Chicago Press, 1986).
7. Davis, M., Efstathiou, G., Frerik, C. S. & White, S. D. M. *Astrophys. J.* **292**, 371–394 (1985).
8. Frerik, C. D., White, S. D. M., Efstathiou, G. & Davis, M. *Nature* **317**, 595–597 (1985).
9. Kaiser, N. in *Large-Scale Structure and Motions in the Universe* 197–213 (Kluwer, Dordrecht, 1989).
10. Audouze, J. in *Astrophysical Cosmology* 395–425 (Pontifical Academy of Sciences, Vatican City, 1981).
11. Blumenthal, G., Faber, S. M., Primak, J. & Rees, M. *Nature* **311**, 517 (1985).
12. Rubin, V. C., Ford, W. K. Jr, Thonnard, N., Roberts, M. S. & Graham, J. A. *Astrophys. J.* **81**, 687 (1976).
13. Smoot, G. F., Gorenstein, M. V. & Muller, R. A. *Phys. Rev. Lett.* **39**, 898 (1977).
14. Cheng, E. S., Saulson, P. R., Wilkenson, D. T. & Corey, B. E. *Astrophys. J.* **232**, L139 (1979).
15. Smoot, G. & Lubin, P. M. *Astrophys. J.* **234**, L83 (1979).
16. Fixsen, D. J., Cheng, E. S. & Wilkenson, D. T. *Phys. Rev. Lett.* **50**, 620 (1983).
17. Tully, R. B. & Fisher, R. 1977, *Astr. Ap.* **54**, 661 (1977).
18. Aaronson, M., Huchra, J., Mould, J., Schechter, P. L. & Tully, R. B. *Astrophys. J.* **258**, 64–76 (1982).
19. Faber, S. M. & Jackson, R. E. *Astrophys. J.* **204**, 668 (1976).
20. Dressler, A. *Astrophys. J.* **281**, 512 (1984).
21. Faber, S. M. et al. *Astrophys. J. Suppl. Ser.* **69**, 763–808 (1989).
22. Davies, R. L. et al. *Astrophys. J. Suppl. Ser.* **64**, 581–600 (1987).
23. Burstein, D. et al. *Astrophys. J. Suppl. Ser.* **64**, 601–642 (1987).
24. Dressler, A. et al. *Astrophys. J.* **313**, 42–58 (1987).
25. Djorgovski, S. & Davis, M. *Astrophys. J.* **313**, 59 (1987).
26. Dressler, A. et al. *Astrophys. J.* **313**, L37–L42 (1987).
27. Lilje, P., Yahil, A. & Jones, B. J. T. *Astrophys. J.* **307**, 91 (1986).
28. Lynden-Bell, D. et al. *Astrophys. J.* **326**, 19–49 (1988).
29. Bertschinger, E. & Juskiewicz, R. *Astrophys. J.* **334**, L59 (1988).
30. Dressler, A. *Astrophys. J.* **329**, 519–526 (1988).
31. Dressler, A. *Astrophys. J. Suppl. Ser.* (in the press).
32. Aaronson, A. et al. *Astrophys. J.* **338**, 654 (1989).
33. Lucey, J. R. & Carter, D. *Mon. Not. R. astr. Soc.* **235**, 1177 (1988).
34. Dressler, A. & Faber, S. M. *Astrophys. J.* **354**, 13–17 (1990).
35. Dressler, A. & Faber, S. M. *Astrophys. J.* **354**, L45–L48 (1990).
36. Burstein, D., Faber, S. M. & Dressler, A. *Astrophys. J.* **354**, 18–32 (1990).
37. Rowan-Robinson, M. et al. *Mon. Not. R. astr. Soc.* **247** (1990).
38. Strauss, M. A. & Davis, M. in *Large Scale Motions in the Universe: A Vatican Study Week* 255–274 (Princeton University Press, New Jersey, 1989).
39. Rowan-Robinson, M. in *Large-Scale Structure and Motions in the Universe* 19–39 (Kluwer, Dordrecht, 1989).
40. Bertschinger, E. & Dekel, A. *Astrophys. J.* **336**, L5 (1989).
41. Willick, J. A. *Astrophys. J.* **351**, L5–8 (1990).
42. Scaramella, R., Baiesi-Pillastrini, G., Chincarini, G., Vettolani, G. & Zamorani, G. *Nature* **338**, 562 (1989).
43. Raychaudhuri, S. *Nature* **342**, 251–255 (1989).
44. Broadhurst, T. J., Ellis, R. S., Koo, D. C. & Szalay, A. S. *Nature* **343**, 726 (1990).
45. Peebles, P. J. E. & Silk, J. *Nature* **346**, 233–239 (1990).
46. Partridge, R. B. in *Astronomy, Cosmology, and Fundamental Physics* 105–130 (Kluwer, Dordrecht, 1989).
47. Readhead, A. C. S. et al. *Astrophys. J.* **346**, 566–587 (1989).
48. Meinhold, P. & Lubin, P. *Astrophys. J.* (in the press).
49. Hill, C. T., Schramm, D. N. & Fry, J. N. *Comm. nucl. Part. Phys.* **19**, 25–39 (1989).
50. Pierce, M. J. & Tully, R. B. *Astrophys. J.* **330**, 579 (1988).

ACKNOWLEDGEMENTS. The maps in Fig. 5 were provided by E. Bertschinger, A. Dekel, M. A. Strauss and A. Yahil.

Primary structure and functional expression from complementary DNA of a brain calcium channel

Yasuo Mori*, Thomas Friedrich*†, Man-Suk Kim*, Atsushi Mikami*, Junichi Nakai*, Peter Ruth‡, Eva Bosse‡, Franz Hofmann‡, Veit Flockerzi‡, Teiichi Furuichi§, Katsuhiko Mikoshiba§||, Keiji Imoto*, Tsutomu Tanabe* & Shosaku Numa*¶

* Departments of Medical Chemistry and Molecular Genetics, Kyoto University Faculty of Medicine, Kyoto 606, Japan

‡ Institut für Medizinische Biochemie, Medizinische Fakultät, Universität des Saarlandes, D-6650 Homburg/Saar, Germany

§ Division of Behavior and Neurobiology, National Institute for Basic Biology, Okazaki 444, Japan

|| Division of Regulation of Macromolecular Function, Institute for Protein Research, Osaka University, Suita 565, Japan

The primary structure of a voltage-dependent calcium channel from rabbit brain has been deduced by cloning and sequencing the complementary DNA. Calcium channel activity expressed from the cDNA is dramatically increased by coexpression of the α_2 and β subunits, known to be associated with the dihydropyridine receptor. This channel is a high voltage-activated calcium channel that is insensitive both to nifedipine and to ω -conotoxin. We suggest that it is expressed predominantly in cerebellar Purkinje cells and granule cells.

VOLTAGE-dependent calcium channels play important roles in the regulation of a variety of cellular functions, including membrane excitability, muscle contraction and synaptic transmission and other forms of secretion. Several types of calcium channel have been distinguished by their electrophysiological and pharmacological properties¹. Complementary DNA cloning studies have previously revealed the primary structures of the dihydropyridine (DHP) receptors from skeletal muscle², heart³ and smooth muscle^{4,5}. Expression of the cDNAs yields DHP-sensitive L-type calcium channels^{3,4,6-8}. Molecular identification of other types of calcium channel, however, has not yet been achieved.

We report here the complete amino-acid sequence of a rabbit brain calcium channel (designated as BI) deduced from the cDNA sequence. Calcium channel activity produced in *Xenopus* oocytes by injection of messenger RNA derived from the cDNA is enormously increased by coinjection of mRNAs specific for the α_2 and β subunits⁹, known to be associated with the DHP receptor. The BI channel expressed is a high voltage-activated calcium channel that is insensitive both to nifedipine, a DHP antagonist, and to ω -conotoxin (ω -CgTx). Blot hybridization analysis of RNA from different rabbit tissues suggests that the BI calcium channel is distributed widely in the brain, being abundant in the cerebellum. Analysis of RNA from the cerebella of mutant mice with different types of cerebellar degeneration suggests that this channel is expressed in Purkinje cells and granule cells.

Protein sequence

Figure 1 shows the primary structure of the rabbit BI calcium channel deduced from the nucleotide sequence of the cloned cDNA (for cloning procedures see legend to Fig. 1); the open

reading frame corresponding to the amino-acid sequences of the skeletal muscle and cardiac calcium channels^{2,3} was adopted and the translational initiation site was assigned to the first ATG triplet that appears downstream of nonsense codons found in-frame. An inserted sequence of 423 nucleotide residues which contains in-frame translational termination codons is found in some clones (λ CBA73 and λ CBP401), whereas it is missing in other clones (λ CBA79 and λ CBP407). This insertion-deletion, which probably results from alternative RNA splicing, gives rise to distinct mRNAs encoding two isoforms of the BI calcium channel (BI-1 and BI-2). These isoforms differ from each other in the carboxy-terminal sequence beginning with amino-acid residue 2,230. BI-1 and BI-2 are composed of 2,273 and 2,424 amino-acid residues, respectively, their relative molecular masses being 257,337 and 273,217 (calculated for the sequences presented in Fig. 1 in the case of predicted amino-acid differences; see legend to Fig. 1). S₁ nuclease analysis suggests that comparable amounts of BI-1 and BI-2 mRNAs are present in rabbit brain (see legend to Fig. 1).

The BI calcium channel shares general structural features with DHP-sensitive calcium channels²⁻⁵ and sodium channels¹⁰, thus apparently having the same transmembrane topology as proposed for the latter voltage-gated channels. The BI channel contains four repeating units of homology and each repeat has one positively charged segment (S₄), which probably represents a voltage-sensing region¹¹, and five hydrophobic segments (S₁, S₂, S₃, S₅ and S₆) (refs 2-5, 10). The conserved charged residues in segments S₂ and S₃ (refs 2-5, 10) are also retained. Amino-acid sequence comparison with other calcium channels reveals 42%, 41% and 42% identities between the BI-1/cardiac, BI-2/cardiac, BI-1/skeletal and BI-2/skeletal pairs, respectively. The regions corresponding to the four internal repeats are relatively well conserved, whereas the remaining regions, all of which are assigned to the cytoplasmic side of the membrane, are less well conserved. The BI channel has two potential N-glycosylation sites¹² assigned to the extracellular side (Asn residues 283 and 1,665), and one of them (residue Asn 283) is conserved in DHP-sensitive calcium channels²⁻⁵. All the potential cyclic AMP-dependent phosphorylation sites¹³ found in the BI channel (residues 95, 425, 431, 1,043 and 2,134 in BI-1 and, in addition, residue 2,407 in BI-2) are assigned to the cytoplasmic side.

Expression of cDNA

Messenger RNA specific for each of the BI calcium channel isoforms, synthesized by transcription *in vitro* of the respective cDNAs, was injected into *Xenopus* oocytes. The injected oocytes were tested for whole-cell current elicited by depolarizing pulses in saline in which Ca²⁺ was replaced by Ba²⁺ (40 mM) and Cl⁻ by methanesulphonate¹⁴. In this solution, endogenous chloride current activated by Ca²⁺ entering the cell was eliminated^{3,14}

† Present address: BASF AG, ZHB/P A30, D-6700 Ludwigshafen, Germany.

¶ To whom correspondence should be addressed.

and inward barium current (I_{Ba}) was observed. Oocytes injected with the BI-1-specific mRNA alone or with the BI-2-specific mRNA alone showed only small I_{Ba} (Table 1); no such current was detected for 20 non-injected oocytes. We then examined possible effects on I_{Ba} of coinjecting mRNAs specific for the α_2 -, β - and γ -subunit polypeptides⁹, known to be associated with the DHP receptor (Table 1). I_{Ba} , induced by the BI-1-specific or the BI-2-specific mRNA, was increased by two orders of magnitude when the mRNAs specific for the α_2 -, β and γ subunits together were coinjected. Coinjection of the α_2 - and β -subunit-specific mRNAs together was sufficient to cause a similar increase in I_{Ba} . Coinjection of the α_2 -subunit-specific mRNA alone or the β -subunit-specific mRNA alone increased I_{Ba} to much smaller extents, the latter being more effective than the former, whereas coinjection of the γ -subunit-specific mRNA alone exerted no significant effect. I_{Ba} was not detected in control oocytes into which a combination of the α_2 -, β - and γ -subunit-

specific mRNAs (26 oocytes) or of the α_2 - and β -subunit-specific mRNAs (26 oocytes) was injected without the BI-1-specific or the BI-2-specific mRNA.

Channel properties

For further characterization of the BI channel, the BI-1 and BI-2 isoforms were produced by injecting the respective mRNAs in combination with the α_2 - and β -subunit-specific mRNAs. The two isoforms showed no significant differences in channel properties. I_{Ba} was elicited by test pulses to potentials more positive than -20 mV (Fig. 2B) and underwent slow inactivation during a 1-s depolarization (Fig. 2A). Changing the holding potential from -60 mV to -50 mV (for 5 s) caused little inactivation of I_{Ba} and holding at -30 mV (for 5 s) inactivated I_{Ba} by less than 50%. I_{Ba} was blocked by low concentrations of Cd^{2+} ; the calculated half-blocking concentration was $0.5 \pm 0.1 \mu M$ (mean $\pm$ s.d., $n = 5$, n indicating the number of oocytes) for both BI-1 and BI-2. I_{Ba} was not affected by addition of $100 \mu M Ni^{2+}$ ($n = 7$ each for BI-1 and BI-2), which is known to block low voltage-activated T-type calcium channels¹. Nifedipine ($10 \mu M$) exerted no effect on I_{Ba} ($n = 7$ each for BI-1 and BI-2; Fig. 2Aa), but Bay K 8644 ($5 \mu M$) reduced the peak I_{Ba} to $83 \pm 5\%$ ($n = 7$ for BI-1) and to $81 \pm 3\%$ ($n = 7$ for BI-2) of the control value (Fig. 2Ab). Inhibition of neuronal high voltage-activated calcium channels by Bay K 8644 has been reported previously^{15,16}. ω -CgTx GVIA ($10 \mu M$) did not alter I_{Ba} ($n = 5$ each for BI-1 and BI-2). This was the case even at a Ba^{2+} concentration of 5 mM ($n = 5$ each for BI-1 and BI-2; Fig. 2Ac), which excludes possible interference of high concentrations of Ba^{2+} with inhibition by ω -CgTx¹⁷. Crude venom from *Agelenopsis aperta*^{18,19} reduced the peak I_{Ba} to $39 \pm 8\%$ ($n = 5$ for BI-1) and to $36 \pm 4\%$ ($n = 5$ for BI-2) of the control value (Fig. 2Ad). These DHP compounds and toxins caused no evident voltage shift in the current-voltage relationship of I_{Ba} (data not shown). Figure 2C shows a record of single-channel currents from a cell-attached patch. The single-channel slope conductance was 16.4 ± 0.7 pS ($n = 3$) for BI-1 and 16.5 ± 3.5 pS ($n = 3$) for BI-2. The BI-1 and BI-2 calcium channel isoforms produced by injecting the respective mRNAs alone without the α_2 - and β -subunit-specific mRNAs, when studied by whole-cell I_{Ba} measurements, showed kinetic properties and pharmacological sensitivities indistinguishable from those of the BI-1 and BI-2 isoforms produced by coinjecting the α_2 - and β -subunit-specific mRNAs (data not shown).

Tissue location

RNA preparations from different rabbit tissues (Fig. 3a) and from different regions of rabbit brain (Fig. 3b) were subjected to blot hybridization analysis with a BI calcium channel cDNA probe. A major hybridizable RNA species of $\sim 9,400$ nucleotides was found abundantly in the brain (Fig. 3a, lane 3) and at a much lower level in the heart (Fig. 3a, lane 2), whereas no hybridization signal was detected with the skeletal muscle, stomach and kidney RNA preparations (Fig. 3a, lanes 1, 4 and 5). This RNA species differs in size from the rabbit brain and heart RNA species hybridizable with the cardiac DHP receptor cDNA³ ($\sim 8,900$ and $\sim 15,500$ nucleotides in length), which suggests the existence of multiple calcium channel mRNA species in these tissues. The BI mRNA was found in different regions of the brain, its level being much higher in the cerebellum (Fig. 3b, lane 3) than in the cerebrum, midbrain and medullas (Fig. 3b, lanes 1, 2 and 4).

To examine the distribution of BI mRNA expression in the cerebellum, we used mutant mice with different types of cerebellar degeneration (Fig. 3c and d). The wild-type mouse cerebellum (Fig. 3c, lane 2 and Fig. 3d, lane 1) contained two major hybridizable RNA species of $\sim 8,200$ and $\sim 9,200$ nucleotides, which apparently represent the mouse BI calcium channel mRNAs. The concentrations of these RNA species were much reduced in the cerebellum of pcd (for Purkinje cell degeneration)

TABLE 1 Effects of addition of α_2 , β and γ subunits on BI calcium channel activity

Subunit-specific mRNAs coinjected	BI-1		BI-2	
	Number of responsive oocytes/number of oocytes tested	Peak I_{Ba} (nA)	Number of responsive oocytes/number of oocytes tested	Peak I_{Ba} (nA)
None	21/57	31 ± 25	23/45	31 ± 21
α_2	20/32	89 ± 54	22/24	93 ± 70
β	24/24	664 ± 623	21/21	562 ± 335
γ	20/40	31 ± 15	22/40	40 ± 21
α_2 , β	26/26	$6,437 \pm 2,575$	28/28	$6,606 \pm 3,224$
α_2 , γ	25/33	79 ± 40	22/28	86 ± 45
β , γ	20/20	791 ± 749	20/20	611 ± 328
α_2 , β , γ	25/25	$6,866 \pm 3,030$	28/28	$6,611 \pm 3,219$

Oocytes were injected with the BI-1-specific or the BI-2-specific mRNA, together with the skeletal muscle α_2 -, β -, or γ -subunit-specific mRNA or their combinations as indicated, and were then tested for I_{Ba} . The holding potential was -60 mV and the test potential was $+20$ mV. Data for peak I_{Ba} are given as means $\pm$ s.d.; the means have been calculated only for the responsive oocytes.

The pSP72 recombinant plasmid pSPCBI-1 carrying the entire protein-coding sequence of the BI-1 cDNA was constructed using the following fragments (*EcoRI* fragments from the cDNA inserts of recombinant phages had been subcloned into the *EcoRI* site of pBluescript SK(-) (Stratagene)): 2.3-kb *SalI* (on pBluescript)-*HindIII* (2, 210) from λ CBRS, *HindIII* (2, 210)-*SmaI* (3, 414) from λ CBP151, *SmaI* (3, 414)-*BglII* (5, 892) (including the coding sequence for sequence a; see legend to Fig. 1) from λ CBP107, *BglII* (5, 892)-*BstXI* (6, 055) from λ CBP126, *BstXI* (6, 055)-*EcoRI* (8, 523) from λ CA73, ~ 0.52 -kb *EcoRI*-*XbaI* (carrying a poly(dA)-poly(dT) tract) from pR1-2A (ref. 11) and 2.5-kb *XbaI*-*SalI* from pSP72 (restriction endonuclease sites are identified by numbers indicating the 5'-terminal nucleotide generated by cleavage). The pSP72 recombinant plasmid pSPCBI-2 carrying the entire protein-coding sequence of the BI-2 cDNA was constructed using the same fragments as described above, except that the *BstXI* (6, 055)-*EcoRI* (8, 523) fragment from λ CA73 was replaced by the *BstXI* (6, 055)-*EcoRI* (8, 523) fragment (with a deletion of residues 6,687-7,109; see legend to Fig. 1) from λ CA79. The pSP65 recombinant plasmid pCAB1 carrying the entire protein-coding sequence of the β -subunit cDNA was constructed using the following fragments: *EcoRI* (linker attached to *HaeIII* (-21))-*RsrII* (120) from pCaChB8-III (ref. 28), *RsrII* (120)-*SacI* (696) from pCaChB11-I (ref. 28), *SacI* (696)-*EcoRI* (linker attached to *PfI* (1,609)) from pCaChB11-II (ref. 28) and *EcoRI*-cleaved pSP65 (containing the ~ 0.26 -kb *BglII*-*BamHI* fragment carrying a poly(dA)-poly(dT) tract from pSPCA1 (ref. 3) at the *BamHI* site). The pSP72 recombinant plasmid pCAG1 carrying the entire protein-coding sequence of the γ -subunit cDNA was constructed by inserting the *NarI* (-47)-*NarI* (681) fragment from the γ -subunit cDNA³⁹ into the *AclI* site of pSP72 (containing the ~ 0.26 -kb *BglII*-*BamHI* fragment from pSPCA1 at the *BamHI* site). mRNAs specific for the BI-1 and BI-2 calcium channel isoforms and for the skeletal muscle α_2 -, β and γ subunits were synthesized *in vitro*⁹ using *XbaI*-cleaved pSPCBI-1, *XbaI*-cleaved pSPCBI-2, *SalI*-cleaved pSPCA1 (ref. 3), *XbaI*-cleaved pCAB1 and *SmaI*-cleaved pCAG1, respectively, as templates. After removal of the follicular cell layer, *Xenopus* oocytes were injected with either the BI-1-specific mRNA ($0.3 \mu g \mu l^{-1}$) or the BI-2-specific mRNA ($0.3 \mu g \mu l^{-1}$) in combination with the α_2 -subunit-specific mRNA ($0.3 \mu g \mu l^{-1}$), the β -subunit-specific mRNA ($0.1 \mu g \mu l^{-1}$) and/or the γ -subunit-specific mRNA ($0.1 \mu g \mu l^{-1}$); the average volume injected was ~ 50 nl per oocyte. The injected oocytes were incubated for 2-3 days as described previously³ and then subjected to electrophysiological measurements. Whole-cell currents were recorded at $20 \pm 2^\circ C$ with a two-microelectrode voltage clamp in a solution of the following composition (in mM): $40 Ba^{2+}$, $50 Na^+$, $2 K^+$ and 5 HEPES (pH 7.4 with methanesulphonic acid). Current records were sampled at 2-ms intervals after low-pass filtering at 0.3 kHz. Leakage currents were eliminated by subtracting the residual currents obtained in the presence of $50 \mu M CdCl_2$.

B1	M	LRALVOPATPAYGLPSHLSAETESTCKGTVVHEAQLNHFYISPGGSNYGSPRAH-ANNANAAAGLAPEHPTTGAALSWQAAIDAAOAKLMGSAG	43
C			99
Sk			1
B1		AORMYKOSMAORARTMALYNPIPVRONCLTVNRSIFLSEDAVVRKAKITTEWPPFEYMTLATIANDQVLAEOHLPDOKTPMSEKDOTEPYFIFGII	143
C		NATISVTSSTOKRROOYQPKKGGSTATRPRAALLCLTKNPIRACISIVIEWKPFETITITITIFANQWALAIYIPFEDDSNATNSNLERVEYFLITIT	199
Sk		EPSSPDDEGLRKKPKPLPEVLPRAALLFCITLQMLRACISIVIEWKPFETITITITIFANQWALAIYIPFEDDSNATNSNLERVEYFLITIT	96
B1		FCFAGIKITIALGFAFKSYLRNGNVMDFVVLITGLIAW	225
C		FTVEAFKVIAMGLFHPNAYLRNGNVLDFITIVVGLFSAILLEQATKADGAN-ALGGKAGFQVKALRAFRVLRPLVSGVPSLOVVUNSTIKAMVF	297
Sk		ESTLAKMKITAMGLFHPNAYLRNGNVLDFITIVVGLFSAILLEQATKADGAN-ALGGKAGFQVKALRAFRVLRPLVSGVPSLOVVUNSTIKAMVF	195
B1		LLOFGLLUFPAITFATIGLEFYMGRFHTTFCEDTDDI-OGES-PAPGTEEPARTCPNGTRGPHYEGNNGITOFDNLIAVILTVFOCITMEGW	320
C		LHIALLVLFVITIIYAIIGLEFYMGRMHTCYNOGGAADPAEDD-PSPCALETGKGRDQ-NGTVCKPGWQGHKGTINDFNFAJAVILTVFOCITMEGW	395
Sk		LHIALLVLFVITIIYAIIGLEFYMGRMHTCYNOGGAADPAEDD-PSPCALETGKGRDQ-NGTVCKPGWQGHKGTINDFNFAJAVILTVFOCITMEGW	294
B1		TOLVYNWDASGNTWMLVFTITITIGSFFMLNLVGLVSGFAKERGRVENRRAPLKLRRDGOITRETMGYMWTSTKAEIWLAEDETDVORHPEDGA	420
C		TOMLVYMDAMGYELPMVYFVSVLIFGSFFMLNLVGLVSGFAKERGRVENRRAPLKLRRDGOITRETMGYMWTSTKAEIWLAEDETDVORHPEDGA	491
Sk		TOMLVYMDAMGYELPMVYFVSVLIFGSFFMLNLVGLVSGFAKERGRVENRRAPLKLRRDGOITRETMGYMWTSTKAEIWLAEDETDVORHPEDGA	391
B1		LRRATIKKSTTDLHPPEEADOLATISVGSPPARASIKSAKLENSFFHKERRMRYFIRMVKTTOAFYMTUSLVALNTLCAVILTVFOCITMEGW	520
C		RNMSPTSETESVNTENVAGG-DIEGECG-ARLAHRISKSFYRWRNRRFORRCKRAAVKSNVYFMYLILFVLVFLNTLIVASLYNORHMLTEVOD	587
Sk		QSDTESQY-IELEGIN-KIT-OFIRHNRWNRVFBKCKDOLVSRVEYMWILVILVFLVFLNTLIVASLYNORHMLTEVOD	465
B1		YAEFIFIGLFMSFEMFKMYGLQTRPYFHSFNCFOGVITIGSTFEVITWAVIKPGTSGFISVLRALRLRIFKMTKYMASLRLVLSLNSMSISLLEF	620
C		TANKALLALFTAEMLKMYSLGLOAYFVSUFNRFOCFVCGGILLETILVETKVMSPGTSVLRVRLRLRIFKMTKYMASLRLVLSLNSMSISLLEF	687
Sk		ANRVLSLLEFEMFKMYGLQTRPYFHSFNCFOGVITIGSTFEVITWAVIKPGTSGFISVLRALRLRIFKMTKYMASLRLVLSLNSMSISLLEF	565
B1		LFLVVFALLGMLFGGDFNFDG-TPPTNFOTIPATIMTVFOITLGEDWNEMVYDGTISOGGVO-GEVVSFEYFVITLFGNYITLNVFLATAVDNLA	718
C		LFLVVFALLGMLFGGDFNFDG-TPPTNFOTIPATIMTVFOITLGEDWNEMVYDGTISOGGVO-GEVVSFEYFVITLFGNYITLNVFLATAVDNLA	787
Sk		LFLVVFALLGMLFGGDFNFDG-TPPTNFOTIPATIMTVFOITLGEDWNEMVYDGTISOGGVO-GEVVSFEYFVITLFGNYITLNVFLATAVDNLA	665
B1		MADELTIK-DEOEELAVNORLALOKAKYEAESVPLSAANNSTAMK-EDOROKPAKSVWEORTSEMRKNLLASREALYSEMDPEERKASYARHLRPD	815
C		MADELTIK-DEOEELAVNORLALOKAKYEAESVPLSAANNSTAMK-EDOROKPAKSVWEORTSEMRKNLLASREALYSEMDPEERKASYARHLRPD	830
Sk		MADELTIK-DEOEELAVNORLALOKAKYEAESVPLSAANNSTAMK-EDOROKPAKSVWEORTSEMRKNLLASREALYSEMDPEERKASYARHLRPD	708
B1		MKTHLDRPLVDPDENNRNNTKSRVAEPTVDORLGOORAEOLRQARHHRDRARPSAHAAAGLDARRPWAGSDEALSREGPYGRESOHOAREGGLP	915
C			
Sk			
B1		PGFWEGERAERKAGDPHRRHARHROGVGGSGSGSPRTGTADGEPRHRRVHRPPGEDGPDOKAERRGRHREGSRPARSGEGEAEGPDGGGGGGGERRRR	1015
C			
Sk			
B1		HRHGPPPAYDOPARRDRERRRRRRTKOTGSGVPVSGPNLSITTRP100DLSOEPLAEDMDNKLNSRLATAEPVSPHENLSHAGLPOSPAKMGSTDP	1115
C			
Sk			
B1		GTPATAANPONSTASRRTPNPNPGNSNPGPKTPENSLLVTNPSTADTNS-IAKARKPDHTTVEIPACDPLNHTVYOVNKNWOPDIPKKEDEKKEE	1214
C		EELTELKSTADGESPPRTIT-KINMODLOPNESEOKSPYPNPETGEDEEEPEMPVGP-RPRP	891
Sk		PMGEGLPTIAKLVDEFSNVNEVKDYPSPADFPDGOGEDEPEIPEVSP-RPRP	760
B1		VOEGPGEOPKPMPPVYSMFITSTIPNRLRLCHYTUNLRYFEMCITLMTVAMSIALAAEDPVPNAPRNNVLYFQVYFTGVFTFEMVYKMLDGLGLH	1314
C		LSELHLKEKAPMPEASAFITSPNNRFRLOCHRIIVNDITITNLLFFITLISLAAEDPVDHTSFNKLILFYFDIVITITITLALKMTAYGALFK	991
Sk		GAELCOKKAVPTEDEASSFFISPTNKVAVLCHRIIVNAWETNELLFFITLISLAAEDPVRAGSVRNQILGYFQIAETISVFTVITLILKMTTYGAFLEK	860
B1		GAVERDPLNITLDFIVVSALVAFATGNSKGDINTKSLRVLRLPLKTIKRLPKKAVFQCVVNSLKWVNTLITVIMDFMFIAMVAVOLFCKGKFFH	1414
C		GSFCRNYNTLDFIVVSALVAFATGNSKGDINTKSLRVLRLPLKTIKRLPKKAVFQCVVNSLKWVNTLITVIMDFMFIAMVAVOLFCKGKFFH	1087
Sk		GSFCRNYNTLDFIVVSALVAFATGNSKGDINTKSLRVLRLPLKTIKRLPKKAVFQCVVNSLKWVNTLITVIMDFMFIAMVAVOLFCKGKFFH	956
B1		CTDESKFEKDKCRKYLITKENEK-ARDREKKEYEFDHNVNLWALLTFTVSTGEWPOVLKHSVDAFENOGGSPGYRMEMSTFYVYVVFVFF	1511
C		CTDESKFEKDKCRKYLITKENEK-ARDREKKEYEFDHNVNLWALLTFTVSTGEWPOVLKHSVDAFENOGGSPGYRMEMSTFYVYVVFVFF	1187
Sk		CTDESKFEKDKCRKYLITKENEK-ARDREKKEYEFDHNVNLWALLTFTVSTGEWPOVLKHSVDAFENOGGSPGYRMEMSTFYVYVVFVFF	1056
B1		VNIFVALITITFOEOGDKMMCEYCLEKNERAETDFAISAKPLTTHMPDOKSOFVMDQFVVPPEYITIMAMALNTIVDMKFKGAV-AYDOKLVF	1610
C		VNIFVALITITFOEOGDKMMCEYCLEKNERAETDFAISAKPLTTHMPDOKSOFVMDQFVVPPEYITIMAMALNTIVDMKFKGAV-AYDOKLVF	1284
Sk		VNIFVALITITFOEOGDKMMCEYCLEKNERAETDFAISAKPLTTHMPDOKSOFVMDQFVVPPEYITIMAMALNTIVDMKFKGAV-AYDOKLVF	1153
B1		NIVFTSLFSLLECLKVLAFGLNTPDIAWNTIDFWITVLSITITDITVTEFG-PSMNAEENSRLSITIFLRLFRVNRKLLSREGIT	1687
C		NIVFTSLFSLLECLKVLAFGLNTPDIAWNTIDFWITVLSITITDITVTEFG-PSMNAEENSRLSITIFLRLFRVNRKLLSREGIT	1376
Sk		NIVFTSLFSLLECLKVLAFGLNTPDIAWNTIDFWITVLSITITDITVTEFG-PSMNAEENSRLSITIFLRLFRVNRKLLSREGIT	1275
B1		RILCLWTFVOSFALPYVCLLITAMFFIYAVITGMQVFGNTGDMEDOSDEDEFOTIENHNFRTFPAALMLFRSATGEAMHNTILSGRPODKNSGT	1786
C		RILCLWTFVOSFALPYVCLLITAMFFIYAVITGMQVFGNTGDMEDOSDEDEFOTIENHNFRTFPAALMLFRSATGEAMHNTILSGRPODKNSGT	1468
Sk		RILCLWTFVOSFALPYVCLLITAMFFIYAVITGMQVFGNTGDMEDOSDEDEFOTIENHNFRTFPAALMLFRSATGEAMHNTILSGRPODKNSGT	1344
B1		LTP-ECQNEFAVYFVSFIFLCSFLMLNLFVAVIMDNFYLTRDSILGPHHLDVYRVWAEYDPAWGRMLYRDMYAMLRHMPPLIGLKNCPLAR	1881
C		NSTEGET-PCGSSFAVYFVSFIFLCSFLMLNLFVAVIMDNFYLTRDSILGPHHLDVYRVWAEYDPAWGRMLYRDMYAMLRHMPPLIGLKNCPLAR	1567
Sk		LTP-ECQNEFAVYFVSFIFLCSFLMLNLFVAVIMDNFYLTRDSILGPHHLDVYRVWAEYDPAWGRMLYRDMYAMLRHMPPLIGLKNCPLAR	1442
B1		VAKRLRLMDLPWADONTVFNSTLALITRTDLKLAROGADKODPAELRKHMAWPNLSORTLQLVTPHKSOTLTGKRLVAMHTMEYRROSKAR	1981
C		VAKRLRLMDLPWADONTVFNSTLALITRTDLKLAROGADKODPAELRKHMAWPNLSORTLQLVTPHKSOTLTGKRLVAMHTMEYRROSKAR	1634
Sk		VAKRLRLMDLPWADONTVFNSTLALITRTDLKLAROGADKODPAELRKHMAWPNLSORTLQLVTPHKSOTLTGKRLVAMHTMEYRROSKAR	1539
B1		KLQ-AMREONRTPIMFORMEPDDEGAGON-ALPSTOLDPAGGLMAHEOGLKQ-SPS-WVTORADEMFOK	2049
C		KEOGLVKGPSORMALSLQAGRLTHD-IGPEIRRAISGDLTAEELDKAMEKVASA-SEDDIFRAGGLFGNHVSYSDSSRSAPHTFTTIRPLHISKA	2763
Sk		QEL-YYGRPKKOTVOTAGLRTIEGALPGRRTISGDLTAEELDERAMEA-AMEERFRRTGLFGGYDTIG-RTNSLHPYMANRPLFOAEI	1633
B1		TGTWSPERAPPADMAOSKOPPOSVEMREMSODGYSOSCHLPMGEOARASMPRLAENORRRGRPRGSDLTICOTSPMKRSASVLGPKASRLDODSL	2149
C		GNNGOTESPHEKLVSTFTSPSSYSTGSANINANN-NALGRLPAPAGYPTSTVEGH-GSPLSPAVRAGEAANKSSKRCHSOESQIAME	1857
Sk		EMEE-LESM-VFLEODPADARTNPLARANTNANANVAYGN-SNHSNNOMFSYVCE	1687
B1-1		ERVPEENORHPPRRERARHTSERSL-GRYTDVDTGLGDLSTMTTOSGDLPSRREOERG-RPKDKRHPHHHHHHHHHPPGRGPRVS	2229
C		QEGASODDN-YDVRIGEDAECCEPSSLSTEMLSYODDENROLAPPEEKDIRLSPKGLFRSASLGRRASFHLECLKROKNOGGDISQKTVLPCHVH	2236
Sk			1956
B1-1		PGVSARRRRRGPVAVRPARAPAL-AHARARAR-PARLLP-ELRLRR-ARRPRRORRRRRRRGGGGRALRRAPGPRELAQDSPPGRGPSVCLARAA	2332
C		HOALAVAGLSPLLORSH-SPTSLPRPCATPPATP-GSRGP-POIPT-LRLEGADSEK-LNSFSPHCGSWGSENSPERG-DSSAARRAR	2043
Sk		EAEETPAAGRGAL-SHSHRALGPHSKPCAGKINGLOVOPGMPINOAPPAPCOOPSTDPPEP-GORRTSLTGLSODEAPORRSSSEGSTRPRA	1781
B1-1		PAGPORLLPGPRTGOAPRRLPOKPARSVORERRGLV-LSPPPPPEGLAARAHPARTPRGPGDSRRGGRRMTASAKSG-GGPPRASAPSP	2273
C		PYSLTVPQOAGAGROFHGSAS-SLYEAVLISEGLFOADOPKFIEVITTOELADACDLTI-EEMNAADDI-LSGGARSPPNGTLFPV	2424
Sk		PATALLIOEA-LVRGGDLTADAGFVATSOALADACOMEP-EEVEVATTEL-LK-ABESVOG-MASV	2129
B1-1			1845
B1-2			
C			
Sk			
B1-1		NRDRPDRORAGNEODASGACAPGCGGSEALADRRAGVSL	2171
C		P-GSLRRS-SLGSLOVQ-SOETLIPPP	1873

FIG. 1 Amino-acid sequence (in single letter code) of the rabbit brain calcium channel BI (top), deduced from the cDNA sequence, and its alignment with those of the rabbit cardiac³ (C, middle) and the rabbit skeletal muscle DHP-sensitive calcium channel² (Sk, bottom). Nucleotide residues are numbered from the first residue of the ATG initiation triplet and the preceding residues are indicated by negative numbers (totalling 8,820 nucleotides; see Acknowledgements for database entry). The 3'-terminal nucleotide residue 8,527 is not followed by a poly(dA) tract. Amino-acid residues are numbered from the initiating methionine. The numbers of the amino-acid residues at the right-hand end of the individual lines are given. The arrowhead indicates where the BI-1 and BI-2 proteins diverge in sequence; residues 2,150–2,229 of BI-1, which immediately precede the position indicated by the arrowhead, are not displayed. Sets of three (or four after divergence of the BI-1 and BI-2 sequences) identical residues at one position are enclosed with solid lines, and sets of three (or four) identical or conservative residues¹⁰ at one position with broken lines. The putative transmembrane segments S1–S6 in each of the repeats I–IV are shown. The amino-acid differences predicted by nucleotide differences observed among the individual clones are as follows: insertion (λ CBP307 and λ CBS) or deletion (λ CBP315) of Gly at residue 419; Ala (λ CBP151) or Thr (λ CBS) at 877; Ser (λ CBP151 and λ CBP107) or Asn (λ CBS) at 1,104; insertion (λ CBRS, λ CBP151 and λ CBP107, and two other clones not fully sequenced) or deletion (λ CBP103) of residues 772–1,120 (λ CBP107 contains the coding sequence for residues 1,052–1,120); residues 1,857–1,884, referred to as sequence a (λ CBP107 and λ CBP126, and one other clone not fully sequenced), or the sequence HYKDMYSLLRVISPLGLGKKCPHRVAC, referred to as sequence b, at the equivalent positions (λ CBP109 and λ CB101). The latter two differences result probably from alternative RNA splicing. S₁ nuclease mapping³⁶ of rabbit brain poly(A)⁺ RNA suggests that mRNA encoding the BI sequence with the insertion of residues 772–1,120 is more abundant than mRNA encoding the BI sequence without this insertion and that all four kinds of mRNA species encoding BI-1 with the sequence a or b and BI-2 with the sequence a or b are present in comparable amounts (for data, see Methods). The amino-acid sequences of BI-1 and BI-2 are presented as composite sequences containing the insertion of residues 772–1,120 and the sequence a, although our data do not answer the question as to whether this insertion is linked with the choice between the sequences a and b and with the insertion–deletion giving rise to BI-1 and BI-2.

METHODS. A randomly primed, size-selected (> ~1 kilobase pair (kb)) cDNA library was constructed³ in phage λ gt10 using poly(A)⁺ RNA prepared⁶ from adult rabbit brain. It was screened with the rabbit skeletal muscle DHP receptor cDNA probe described previously³ to yield clone λ CB3 (3,727–5,050); numbers in parentheses indicate the nucleotide residues of the cDNA carried by the clone. Restriction fragments from λ CB3 were used as probes to clone adjacent cDNAs and similar cloning procedures were repeated. λ CBS (–34 to 3,394) and λ CB101 (5,321–6,227) were isolated from another randomly primed, size-selected (> ~0.7 kb) cDNA library, and λ CBA73 (5,958–8,527) and λ CBA79 (5,927–8,527 with a deletion of 6,687–7,109; some nucleotide differences are found in the 3'-noncoding region as compared with λ CBA73) from an oligo(dT)-primed, size-selected (> ~2 kb) cDNA library. Three synthetic primers complementary to nucleotide residues 7,793–7,809 (primer I), 3,812–3,828 (primer II) and 1,860–1,876 (primer III) were elongated and cloned into λ gt10 by the procedures described previously³ (size selection > ~0.5 kb for primer I; > ~0.7 kb for primer II; > ~1.0 kb for primer III). λ CBP401 (6,047–7,809) and λ CBP407 (5,947–7,808 with a deletion of 6,687–7,109) were selected from the primer I-derived cDNAs, clones λ CBP103 (1,777–3,827 with a deletion of 2,315–3,361), λ CBP151 (1,858–3,819), λ CBP107 (3,152–5,898), λ CBP109 (3,639–6,313) and λ CBP126 (5,224–6,305) from the primer II-derived cDNAs, and λ CBP315 (–555 to 1,866 with a deletion of 1,255–1,257) and λ CBP307 (–293 to 1,866) from the primer III-derived cDNAs. Appropriate restriction fragments from the isolated clones were subcloned into pBluescript SK (–). Nested deletions were made³ and DNA sequencing³ was carried out on both strands. The probes used for S₁ nuclease mapping and the sizes (in nucleotides) of protected fragments were as follows: ~60 and ~160 from the 1.1-kb *Pvu*I (on pSP72)–**Sma*I (3,414) fragment containing nucleotides 2,209–3,413 with a deletion of 2,315–3,361 (the site of 5'-end labelling is marked with an asterisk); ~60 and ~160 from the 1.1-kb *Pvu*I (on pSP72)–**Sma*I (3,414) fragment containing 3,250–3,413; ~1,500 and ~1,600 (in addition ~300 and ~800) from the 2.6-kb *Pvu*I (on pSP73)–**Hinf*I (7,125) fragment containing 5,486–7,124 (including the coding sequence for sequence a); ~1,100 and ~1,200 (in addition ~340) from the 2.1-kb *Pvu*I (on pSP73)–**Hinf*I (7,125) fragment containing 5,486–7,124 (including the coding sequence for sequence a) with a deletion of 6,687–7,109; ~140 and ~900 (in addition ~420) from the 1.8-kb *Pvu*I (on pSP72)–**Sac*I (7,248) fragment containing 6,355–7,247; ~140 and ~470 from the 1.4-kb *Pvu*I (on pSP72)–**Sac*I (7,248) fragment containing 6,355–7,247 with a deletion of 6,687–7,109.

mice (Fig. 3c, lane 3), where virtually all Purkinje cells are lost²⁰. This suggests that the BI mRNAs are expressed abundantly in cerebellar Purkinje cells. The concentrations of the BI mRNAs were also decreased in the cerebellum of weaver mice (Fig. 3c, lane 1), where almost all granule cells are lost²¹. If the BI mRNAs were present exclusively in Purkinje cells, the weaver cerebellum should show an increase in BI mRNA concentration because of its increased relative content in Purkinje cells, as is the case for the mRNA encoding the inositol 1,4,5-trisphosphate-binding protein P₄₀₀ (ref. 22). Thus our data suggest that the BI mRNAs are also expressed in granule cells. The strong reduction in BI mRNA levels observed in the cerebellum of *pcd* mice may be ascribed partly to secondary consequences of the mutation which result from lack of trophic interactions between Purkinje cells and granule cells²³. The concentrations of the BI mRNAs were also lowered in the cerebellum of staggerer mice (Fig. 3d, lane 2), where Purkinje cells undergo an abortive developmental process²⁴ so that they have stunted dendrites²⁵ and fail to fire Ca²⁺ spikes²⁶. This may indicate that the BI calcium channel is localized in the dendritic membranes of Purkinje cells.

Discussion

Our results indicate that the BI protein alone is sufficient to exhibit calcium channel activity, but that its channel activity is

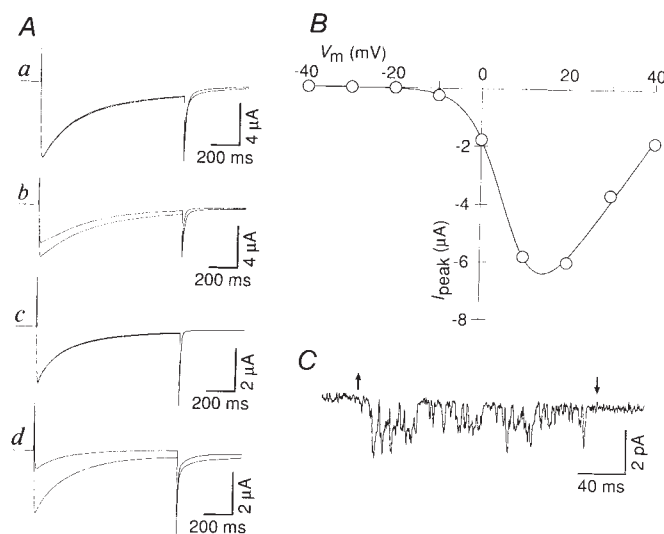


FIG. 2 Properties of I_{Ba} in *Xenopus* oocytes injected with the mRNA specific for the BI-2 calcium channel in combination with the mRNAs specific for the skeletal muscle α_2 and β subunits. **A**, Whole-cell currents elicited by a 1-s test pulse from -60 mV to +20 mV before and during exposure to 10 μM nifedipine (**a**, overlapping the control trace), 5 μM Bay K 8644 (**b**, upper trace), 10 μM ω-CgTx GVIA (**c**, overlapping the control trace) or *A. aperta* venom (**d**, upper trace). Leakage currents were not subtracted. **B**, Whole-cell peak I_{Ba} (I_{peak}) plotted against test potential (V_m). Holding potential -60 mV. **C**, Single-channel currents recorded from a cell-attached patch. Inward current downwards. The arrows indicate the beginning and end of a test pulse from -60 mV to +10 mV relative to the resting potential of the oocyte. The maximum inward current in the trace shown corresponds apparently to opening of three channels, one or more of which are in a subconductance state.

METHODS. The experimental conditions used were the same as in Table 1 unless otherwise specified. In the measurements shown in **Ac**, the Ba²⁺ concentration used was 5 mM (see text). Crude venom from *A. aperta* (Spider Pharm) was applied at 1:1,000 dilution. Single-channel current measurements were performed at 20 ± 1 °C on oocytes bathed in a depolarizing solution of the following composition (in mM): 100 KCl, 10 EGTA and 10 HEPES (pH 7.2 with KOH). The pipette solution contained (in mM): 110 BaCl₂ and 10 HEPES (pH 7.4 with Ba(OH)₂). Single-channel currents were filtered at 0.5 kHz and sampled at 0.2-ms intervals. Linear leakage and capacitive currents were eliminated using the P/4 method. Single-channel slope conductance was determined by linear regression of current amplitudes of long, well-resolved openings at test potentials ranging from -5 mV to +50 mV.

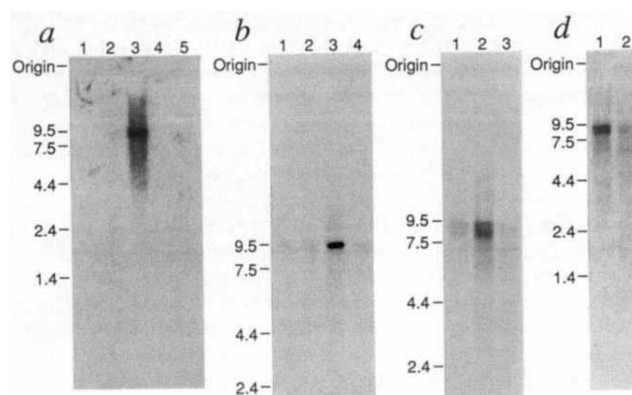


FIG. 3 Autoradiograms of blot hybridization analysis with a BI calcium channel cDNA probe of RNA from different rabbit tissues (a), different regions of rabbit brain (b) and cerebella of normal and mutant mice (c and d). a, Analysis of poly(A)⁺ RNA (10 µg each) prepared⁶ from skeletal muscle (lane 1), heart (lane 2), brain (lane 3), stomach (lane 4) and kidney (lane 5) from adult rabbits. b, Analysis of total RNA (10 µg each) prepared⁶ from cerebellum (lane 1), midbrain (lane 2), cerebellum (lane 3) and medulla-pons (lane 4) from adult rabbits. c, Analysis of total RNA (10 µg each) prepared⁶ from cerebella of 1-month-old wild-type CD-1 (ICR) (lane 2), pcd (lane 3) and weaver mice (lane 1). d, Analysis of total RNA (10 µg each) prepared from cerebella of 1-month-old wild-type CD-1 (ICR) (lane 1) and staggerer mice (lane 2). The procedures used were as described previously³⁷ and the probe was the Bbv1 (3,843)–Bbv1 (5,088) fragment derived from λ CBP107 and labelled by nick translation. Autoradiography was at -70°C for 24 h (a) or 7 days (b–d) with an intensifying screen. An RNA ladder (Bethesda Research Laboratories) was used as size markers (in kilobases).

dramatically increased by coexpression of the skeletal muscle α_2 and β subunits, known to be associated with the DHP receptor. There is evidence that the brain contains proteins homologous (or identical) with the skeletal muscle α_2 and β subunits^{27–29}. Thus it seems likely that different types of voltage-dependent calcium channel in different tissues have a similar subunit structure, as also suggested by immunoprecipitation studies on DHP-sensitive calcium channels²⁹. The mechanism underlying the observed effect of coexpression of the α_2 and β subunits is not known. These subunits may increase expression of the BI calcium channel by stabilizing the protein or facilitating its insertion into the membrane. Consistent with this notion is our finding that coexpression of these subunits does not appreciably affect the macroscopic current properties and pharmacological sensitivities of the BI channel. It remains possible, however, that these subunits may modulate single-channel properties of the BI channel.

Neuronal voltage-dependent calcium channels are commonly classified into three types: low voltage-activated T-type, high voltage-activated DHP-sensitive L-type and high voltage-acti-

vated ω -CgTx-sensitive N-type¹. It has recently been reported, however, that various neurons contain a high voltage-activated calcium channel that is resistant to both DHP and ω -CgTx^{30–32}. This type of channel is a major calcium channel component in cerebellar Purkinje cells^{18,30,31} and granule cells³² and is inhibited by *A. aperta* venom^{18,32}. Thus the electrophysiological and pharmacological properties of the BI channel as well as its presence in Purkinje cells and granule cells suggest that it represents this new type of calcium channel. The reported insensitivity to Bay K 8644 of calcium channels in Purkinje cells³⁰, as opposed to the inhibition observed for the BI calcium channel, may be accounted for by possible activation of minor L-type calcium channels in these cells masking inhibition of the BI calcium channel. Voltage-dependent calcium channels, distributing over the dendritic membranes of Purkinje cells, are capable of generating Ca^{2+} -dependent action potentials³³ which may play an essential role in inducing long-term depression³⁴ in the cerebellar cortex. The BI calcium channel may be involved in this function. □

Received 14 December 1990; accepted 15 February 1991.

1. Tsien, R. W., Lipscombe, D., Madison, D. V., Bley, K. R. & Fox, A. P. *Trends Neurosci.* **11**, 431–438 (1988).
2. Tanabe, T. *et al.* *Nature* **328**, 313–318 (1987).
3. Mikami, A. *et al.* *Nature* **340**, 230–233 (1989).
4. Biel, M. *et al.* *FEBS Lett.* **269**, 409–412 (1990).
5. Koch, W. J., Ellinor, P. T. & Schwartz, A. *J. Biol. Chem.* **265**, 17786–17791 (1990).
6. Tanabe, T., Beam, K. G., Powell, J. A. & Numa, S. *Nature* **336**, 134–139 (1988).
7. Perez-Reyes, E. *et al.* *Nature* **340**, 233–236 (1989).
8. Tanabe, T., Mikami, A., Numa, S. & Beam, K. G. *Nature* **344**, 451–453 (1990).
9. Campbell, K. P., Leung, A. T. & Sharp, A. H. *Trends Neurosci.* **11**, 425–430 (1988).
10. Numa, S. & Noda, M. *Ann. N. Y. Acad. Sci.* **479**, 338–355 (1986).
11. Stühmer, W. *et al.* *Nature* **339**, 597–603 (1989).
12. Bause, E. *Biochem. J.* **209**, 331–336 (1983).
13. Krebs, E. G. & Beavo, J. A. *Rev. Biochem.* **48**, 923–959 (1979).
14. Dascal, N., Snutch, T. P., Lübbers, H., Davidson, N. & Lester, H. A. *Science* **231**, 1147–1150 (1986).
15. Boll, W. & Lux, H. D. *Neurosci. Lett.* **56**, 335–339 (1985).
16. Kasai, H., Aozaki, T. & Fukuda, J. *Neurosci. Res.* **4**, 228–235 (1987).
17. McCleskey, E. W. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 4327–4331 (1987).
18. Llinás, R., Sugimori, M., Lin, J.-W. & Cherksey, B. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1689–1693 (1989).
19. Lin, J.-W., Rudy, B. & Llinás, R. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4538–4542 (1990).
20. Mullen, R. J., Eicher, E. M. & Sidman, R. L. *Proc. natn. Acad. Sci. U.S.A.* **73**, 208–212 (1976).
21. Rakic, P. & Sidman, R. L. *J. comp. Neurol.* **152**, 103–132 (1973).
22. Furuichi, T. *et al.* *Nature* **342**, 32–38 (1989).

23. Purves, D. & Lichtman, J. W. *Principles of Neural Development* (Sinauer, Sunderland, Massachusetts, 1985).
24. Hirano, A. & Dembitzer, H. M. *J. Neuropath. exp. Neurol.* **34**, 1–11 (1975).
25. Bradley, P. & Berry, M. *Brain Res.* **142**, 135–141 (1978).
26. Crepel, F., Dupont, J.-L. & Gardette, R. *J. Physiol., Lond.* **346**, 111–125 (1984).
27. Ellis, S. B. *et al.* *Science* **241**, 1661–1664 (1988).
28. Ruth, P. *et al.* *Science* **245**, 1115–1118 (1989).
29. Ahlman, M. K., Westenbroek, R. E. & Catterall, W. A. *Neuron* **4**, 819–832 (1990).
30. Regan, L. J. *Ann. N. Y. Acad. Sci.* **560**, 121–123 (1989).
31. Sah, D. W. Y., Regan, L. J. & Bean, B. P. *Soc. Neurosci. Abstr.* **15**, 823 (1989).
32. Bertolino, M., Vicini, S., Costa, E. & Llinás, R. *FASEB J. Abstr.* **4**, A1202 (1990).
33. Llinás, R. & Sugimori, M. *J. Physiol., Lond.* **305**, 197–213 (1980).
34. Ito, M. A. *Rev. Neurosci.* **12**, 85–102 (1989).
35. Bosse, E. *et al.* *FEBS Lett.* **267**, 153–156 (1990).
36. Berk, A. J. & Sharp, P. A. *Cell* **12**, 721–732 (1977).
37. Mishina, M. *et al.* *Nature* **321**, 406–411 (1986).

ACKNOWLEDGEMENTS. We thank T. Miyata and N. Iwabe for computer analysis. This investigation was supported in part by the Ministry of Education, Science and Culture of Japan, the Mitsubishi Foundation and the Japanese Foundation of Metabolism and Diseases and by fellowships from the Japan Society for the Promotion of Science and the Feodor Lynen Program of the Alexander von Humboldt-Stiftung (T. Friedrich). The nucleotide sequences of the rabbit BI calcium channel cDNAs will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers X57476 (BI-1) and X57477 (BI-2).

Constraints on r-process nucleosynthesis in accretion disks

Liping Jin*

Department of Astronomy, University of Texas at Austin, Austin, Texas 78712, USA

It has been suggested¹ that matter in accretion disks around compact objects such as black holes or neutron stars, which appear to be responsible for galactic X-ray sources, may experience ideal conditions for classical rapid neutron-capture ('r-process') nucleosynthesis². Systems in which accretion drives an outflow from a region near the compact object might thereby enrich the interstellar medium in r-process elements, such as europium. Here I present a more detailed assessment of the efficacy of this mechanism for the r-process. By considering the constraints imposed by typical accretion-disk conditions, I conclude that r-process elements are unlikely to have been made this way, largely because the total production is too low, by a factor of $\sim 10^5$, to explain the observed abundances.

The existence of thick accretion disks (or tori) of hot gas ($kT \approx 1\text{--}100$ MeV) supported by ion pressure in the vicinity of accreting black holes has been proposed to explain the X-ray emission from Cygnus X-1 (ref. 3) and some types of active galactic nuclei⁴. When the accretion rate is low, the inflow timescale may be shorter than the electron-ion coupling time, causing the infalling gas to heat up without being able to cool. In such situations, the flow pattern would consist of a thin disk at large radii ($r > 100r_g$, where $r_g \equiv GM/c^2$, M being the mass of the central object), and a thick toroidal structure at small radii, supported primarily by the pressure of ions at the virial temperature. Here I refer to these regions as the 'cool disk' and the 'hot torus' respectively. Under the conditions in the torus, the thermal energy of ions can be larger than thresholds of nuclear spallation (for details of physical conditions and spallation in the torus; see refs 5–7) and infalling nuclei (particularly helium) are dissociated into protons and neutrons. Hogan and Applegate¹ suggested that a dense reservoir of neutrons may build up and that, if 'seed' nuclei (iron peak elements) encounter this reservoir before being ejected in a hot coronal wind or jet, the system can act as a source of the r-process elements. Here I point out some severe constraints on this mechanism and discuss its viability.

The viscous stress may be written as αP , where P is the pressure⁸ and α is the viscosity parameter (≤ 1). If α is fixed, the inflow time in a disk of half-thickness $h(r)$ around a compact object is⁴

$$t_i = \alpha^{-1} \left(\frac{r}{r_g} \right)^{3/2} \left(\frac{h}{r} \right)^{-2} \left(\frac{r_g}{c} \right) \quad (1)$$

The condition for electron-ion coupling to be ineffective in the inner parts of a torus is

$$\dot{m} < 50\alpha^2 \quad (2)$$

where $\dot{m}$ is the mass accretion rate in units of the Eddington rate $\dot{M}_{\text{Edd}} \equiv L_{\text{Edd}}/c^2$. Equation (2) is a necessary condition for a hot torus to exist: if this condition (which I call the 'existence condition') is not satisfied, the disk will not be hot enough to dissociate nuclei into neutrons and protons. When equation (2) holds, the ions can remain at the virial temperature

$$\begin{aligned} T_i &\approx T_{\text{virial}} = \left(\frac{m_p c^2}{3k} \right) \left(\frac{r_g}{r} \right) = \left(\frac{315 \text{ MeV}}{k} \right) \left(\frac{r_g}{r} \right) \\ &= 3.6 \times 10^{12} \left(\frac{r_g}{r} \right) \text{K} \end{aligned} \quad (3)$$

Such a disk will be pressure-supported with $h \approx r$. Within a radius $r = 100r_g$, kT_i could be 1–100 MeV. If α is independent of radius, the density in the hot torus is

$$\rho = \frac{3}{2} \frac{m_p}{\sigma_T r_g} \frac{\dot{m}}{\alpha} \left(\frac{r}{r_g} \right)^{-3/2} = 2.6 \times 10^{-5} \frac{\dot{m}}{m\alpha} \left(\frac{r}{r_g} \right)^{-3/2} \text{ g cm}^{-3} \quad (4)$$

where $m \equiv M/M_\odot$. The torus is not in local thermodynamic equilibrium; it is generally optically thin, and the photon density is well below that of the particles.

In a hot accretion disk, neutrons can be produced by spallations between helium nuclei and protons. If all helium nuclei are dissociated into protons and neutrons, the rate at which the infalling matter produces neutrons, $\dot{M}_{\text{supply}}$, will be $\sim 0.1\dot{M}$, but could be much lower than this if $\dot{m}\alpha^{-2}$ is much less than 10 (ref. 7). Once neutrons are liberated, they form an electrically neutral disk or torus which evolves separately from the proton/electron disk. Neutrons experience their own collisional viscosity, which produces a neutron infall rate $\dot{M}_n$. A necessary condition for neutrons to accumulate and form a dense reservoir is that $\dot{M}_{\text{supply}}$ exceeds $\dot{M}_n$, that is, $0.1\dot{m} > \dot{m}_n$ (where $\dot{m}_n \equiv \dot{M}_n/\dot{M}_{\text{Edd}}$). I will refer to this as the 'accumulation condition'. Combining the existence condition and the accumulation condition, we have

$$\alpha^2 > 0.2\dot{m}_n \quad (5)$$

Thus if α is too low, this mechanism for the r-process will not operate. From equation (5) of ref. 1, $\dot{m}_n \geq 0.3$. So the accumulation condition requires that $\dot{m} > 3$ and equation (5) above requires that $\alpha > 0.2$ for r-process nuclei to be produced.

When $\dot{M}_{\text{supply}} > \dot{M}_n$, the mass and hence the density of the neutron torus grows until $\dot{M}_{\text{supply}} \approx \dot{M}_{\text{decay}}$, where $\dot{M}_{\text{decay}}$ is the neutron decay rate. $\dot{M}_n$ corresponds to a local spreading rate. It is important to note that the neutrons spreading to larger radii will cool adiabatically to energies ≤ 1 MeV, at which neutron absorption cross-sections become large. The r-process is likely to occur for $r > 100r_g$. Two criteria must be satisfied to produce r-process elements: sufficiently rapid neutron addition and a sufficient total neutron exposure¹. A seed nucleus captures neutrons at a rate $n\sigma V_{\text{thermal}} \approx 10^{3.5} (n_n/10^{20} \text{ cm}^{-3}) \text{ s}^{-1}$, which is fairly independent of V_{thermal} as long as $kT \leq 1$ MeV. Adding ~ 200 neutrons to build the heaviest r-process elements takes a time $\tau_{\text{nadd}} \approx 0.1 \text{ s} (n_n/10^{20} \text{ cm}^{-3})^{-1}$. The 'rapidity requirement' for the r-process is that $\tau_{\text{nadd}} < \tau_\beta$, where τ_β is the sum of the beta-decay half-lives of elements along the r-process path. This requires a high neutron density—this density is higher at smaller radii. If the r-process happens at $r \approx 100r_g$, then from equation (7) of ref. 1, we have $n_n \approx \dot{m}m^{-2} \times 10^{20.5} \text{ cm}^{-3}$. The rapidity requirement implies

$$m^2 < 10^{0.5} \dot{m} \left(\frac{\tau_\beta}{0.1 \text{ s}} \right) \quad (6)$$

Therefore a small m is required for the r-process to occur via this mechanism. Estimates give⁹ $0.1 \text{ s} \leq \tau_\beta \leq 10 \text{ s}$. Taking the maximum possible value of ~ 50 for $\dot{m}$ (see equation (2)), equation (6) gives $m < 100$ for $\tau_\beta = 10 \text{ s}$, and $m < 10$ for $\tau_\beta = 0.1 \text{ s}$.

For the r-process to occur it is necessary that a seed nucleus be exposed to neutrons for a time $\geq \tau_\beta$. The exposure requirement implies that matter ultimately destined for an outflow must reside in the cool part (≤ 1 MeV) of the neutron torus for at least a time τ_β . Assuming the infall time (at $r \approx 100r_g$) to be equal to the exposure time, we have $t_i > \tau_\beta$. Combining the rapidity requirement in equation (6) and the exposure requirement, a successful r-process requires

$$\frac{\tau_\beta}{0.1 \text{ s}} < 8 \times 10^{-3} \frac{\dot{m}}{\alpha^2} \quad (7)$$

Therefore a short τ_β is required for this mechanism to be effective. Taking the maximum possible value of ~ 50 for $\dot{m}/\alpha^2$, equation (7) implies that $\tau_\beta < 0.04 \text{ s}$, which is lower than the values estimated⁹. So this mechanism will be unlikely to work

* Present address: Department of Physics, University of Illinois, 1110 West Green Street, Urbana, Illinois 61801, USA.

unless the exposure time is greater than t_i , which would seem to be difficult to arrange.

It has been suggested¹ that galactic X-ray sources may be sites of significant r-process activity. The ratio of the calculated abundance from this source to the observed abundance (roughly $10^4 M_\odot$ of r-process elements produced in the 10^{10} -yr galactic history¹) is:

$$R \approx 10^{-5} \left(\frac{\dot{M}}{10^{-9} M_\odot \text{yr}^{-1}} \right) \left(\frac{\bar{N}}{300} \right) \left(\frac{X_s}{10^{-3}} \right) \left(\frac{f_e}{0.1} \right) \left(\frac{f_h}{0.1} \right) \left(\frac{f_i}{1.0} \right) \quad (8)$$

where $\dot{M}$ is the average accretion rate of an X-ray source, $\bar{N}$ is the number of X-ray sources averaged over the galactic lifetime, X_s is the abundance of seed nuclei (iron-peak), f_e is the ratio of matter ejected in the outflow to matter accreted, f_h is the fraction of X-ray sources that have hot disks, and f_i is the fraction of seed nuclei in the outflow that escape spallation and undergo r-processing. I have used optimistic estimates for each of these parameters; it is therefore extremely difficult to see where the additional factor of 10^5 needed to account for the galactic abundance of r-process elements could come from. Thus even if the intrinsic constraints discussed above can be avoided, the production by accretion disks in galactic X-ray sources is too low.

I now discuss the factors in equation (8) in turn. A typical X-ray source accretes at a rate $\dot{M} \approx 10^{-9} M_\odot \text{yr}^{-1}$. The present value of $\bar{N}$ is about 300 (ref. 10). There could have been more objects in the early history of the Galaxy, but given the observed age-abundance relation¹¹, it is not possible that the early r-process rate could be several orders of magnitude higher than the current rate. Note that any undetected sources would have $\dot{M} < 10^{-9} M_\odot \text{yr}^{-1}$; disks with such low accretion rates do not contribute to the r-process because of the neutron accumulation condition (see above). The factor X_s appears in equation (8) because only iron-peak elements can become r-process elements. I have (optimistically) assumed solar abundances for these seed nuclei. A good guess of the ejection fraction f_e is 10% (ref. 12). The standard accretion-disk model for X-ray sources is the thin-disk model (which is not hot enough to make neutrons). Only those sources producing hard X-rays (such as Cyg X-1) are explained by the hot-disk model. Therefore this mechanism for the r-process probably does not operate in most X-ray sources. I have assumed that the fraction of hot disks, f_h , is ~ 0.1 (ref. 13). Seed nuclei (iron-peak nuclei) that enter the hot torus may be destroyed by spallation reactions^{7,14}. Only those seed nuclei that stay at temperatures of $kT \leq 1$ MeV for long enough and then escape can contribute to the r-process. So f_i could be small; nevertheless, I have assumed its maximum value of 1.0 in equation (8). The reason for the difference between my conclusion and that of ref. 1 is that, in the latter work, some of the factors in equation (8) were not considered and the assumed efficiency of r-process-element production was too high. $\square$

Spectral transformation of the unusual variable star MWC560 to resemble a nova

Stephen P. Maran, Andrew G. Michalitsianos, Ronald J. Oliveresen & George Sonneborn

Laboratory for Astronomy and Solar Physics, NASA/Goddard Space Flight Center Greenbelt, Maryland 20771, USA

MWC560 is an emission-line star catalogued¹ in 1943 and later described² as an 'extraordinary symbiotic-like variable'. It was recently found³ to be undergoing a photometric and spectroscopic outburst. A dramatic change has occurred in the ultraviolet spectrum of MWC560, so that it now closely resembles the spectrum of a nova shortly after outburst. This event, detected by the International Ultraviolet Explorer satellite, may signal a major mass-ejection episode such as presumably occurred in past centuries in the symbiotic star R Aquarii to produce the well-known bipolar nebula, and it may herald the emergence of a standard symbiotic-star emission-line spectrum in MWC560, corresponding to a change in evolutionary state.

MWC560 is probably a binary system comprising a spectral type M4e red giant⁴ and a hot compact companion star. The discovery of the 1990 outburst³ brought it to the attention of astronomers worldwide. Monitoring MWC560 with the International Ultraviolet Explorer satellite (IUE) during 1990 February 4–1990 April 29, we found that the ultraviolet spectrum was substantially blue-shifted (with velocities of approach on various occasions ranging from 500 to 5,300 km s⁻¹) and that there are pronounced changes in velocity, absorption line strengths and in ultraviolet and visible-light fluxes which were not obviously correlated with one another⁵. The ultraviolet spectrum was dominated by strong, blended absorption lines and line blends of cool ions such as Fe II and Cr II (Fig. 1), as in 1984 (ref. 2), although in 1990 the star was much brighter and the velocities inferred from spectral lines much higher than in 1984. This unusual ultraviolet spectrum more closely resembles that of the peculiar Be star XX Oph ('Merrill's iron star') than any other with which we are familiar, although the highest velocities observed in MWC560 are about 10 times as great as those found in XX Oph (B. Bopp, personal communication).

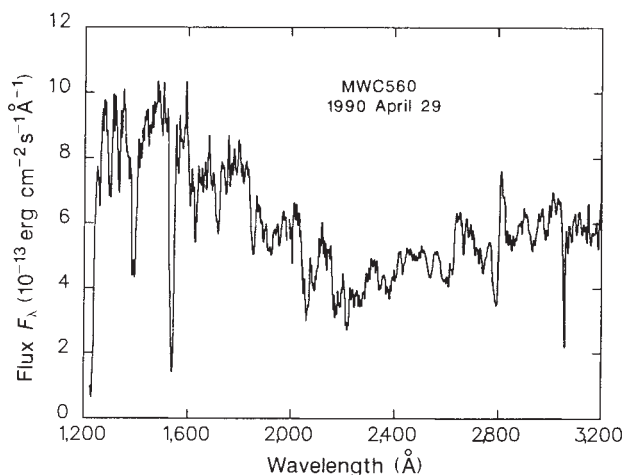


FIG. 1 Ultraviolet spectrum of MWC560 on 1990 April 29 (wavelength range 1,200–3,200 Å) as obtained at low dispersion with the IUE short- and long-wavelength spectrographs. The prominent absorption features are the Fe II blend near 1,384 Å and the Si II line, possibly blended with P II, near 1,535 Å. Note the P Cyg-type line profile of the Mg II feature at 2,800 Å.

Received 9 October 1990; accepted 12 February 1991.

- Hogan, C. J. & Applegate, J. H. *Nature* **330**, 236–238 (1987).
- Burbridge, E. M., Burbridge, G. R., Fowler, W. A. & Hoyle, F. *Rev. mod. Phys.* **29**, 547–650 (1957).
- Shapiro, S. L., Lightman, A. P. & Eardley, D. M. *Astrophys. J.* **204**, 187–199 (1976).
- Rees, M. J., Begelman, M. C., Blandford, R. D. & Phinney, E. S. *Nature* **295**, 17–21 (1982).
- Rees, M. J. in *Formation and Evolution of Galaxies and Large Structures in the Universe* (eds Audouze, J. & Tran Thanh Van, J. 271–277) (Reidel, Dordrecht, 1984).
- Ramadurai, S. & Rees, M. J. *Mon. Not. R. astr. Soc.* **215**, 53P–56P (1985).
- Jin, L. *Astrophys. J.* **356**, 501–506 (1990).
- Shakura, N. I. & Sunyaev, R. A. *Astr. Astrophys.* **24**, 337–355 (1973).
- Schramm, D. N. in *Essays in Nuclear Astrophysics* (eds Barnes, C. A., Clayton, D. D. & Schramm, D. N.) 325–354 (Cambridge University Press, 1982).
- Shapiro, S. L. & Teukolsky, S. A. *Black Holes, White Dwarfs, and Neutron Stars* (Wiley, New York, 1983).
- Mathews, G. J. & Cowan, J. J. *Nature* **345**, 491–494 (1990).
- Lovelace, R. V. E., Wang, J. C. L. & Sulkkanen, M. E. *Astrophys. J.* **315**, 504–535 (1987).
- Eardley, D. M., Lightman, A. P., Payne, D. G. & Shapiro, S. L. *Astrophys. J.* **224**, 53–61 (1978).
- Aharonian, F. A. & Sunyaev, R. A. *Mon. Not. R. astr. Soc.* **210**, 257–277 (1984).

ACKNOWLEDGEMENTS. I thank R. Duncan and E. Vishniac for their critical reading of the manuscript and helpful discussion. I am grateful to J. Truran for helpful discussion. This research was supported in part by an NSF grant to the University of Texas at Austin and in part by NSF and NASA grants to the University of Illinois.

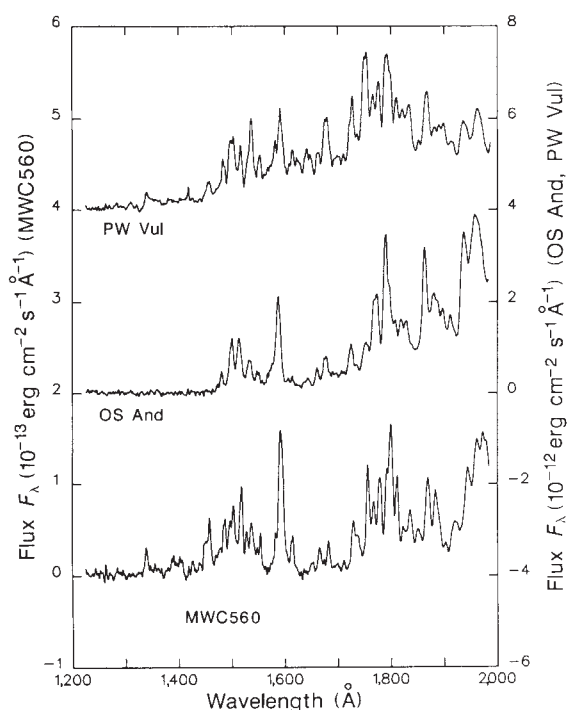


FIG. 2 Low-dispersion ultraviolet spectra (range 1,200–2,000 Å) of MWC560 (bottom) on 1990 September 26, nova OS And (centre), three days after maximum visible light, and nova PW Vul (top), seven days after maximum, obtained with the IUE short-wavelength spectrograph. Note the strong resemblances. The left-hand ordinate scale applies to MWC560, the right-hand scale to the two novae. For purposes of presentation, the PW Vul flux has been multiplied by 4 and then increased by $4 \times 10^{-12} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ Å}^{-1}$.

When we were next able to observe MWC560 with the IUE, on 1990 September 26, we discovered that the ultraviolet spectrum no longer resembled that just described, but now closely resembled that of a classical nova in outburst several days after reaching maximum visible light. The form of the nova spectrum at that time is attributed to an optically-thick shell or 'pseudo-photosphere'. Figures 2 and 3 compare two regions of the ultraviolet spectrum of MWC560 with the corresponding spectra of Nova Andromedae 1986 (OS And), as recorded by the IUE on 1986 December 11, three days after maximum visible light, and Nova Vulpeculae 1984 No. 1 (PW Vul) on 1984 September 1, seven days after maximum⁷. We determined dates of maxima (1984 August 5 for PW Vul, 1986 December 9 for OS And) from photometry reported in refs 8, 9 and 10, and from visible-light photometry with the IUE Fine Error Sensor archived at the IUE Regional Data Analysis Facility at Goddard Space Flight Center. The close correspondence between the large- and small-scale structures in these spectra points to common physical processes. The highly-structured spectra of novae in this phase are thought to represent the residual flux escaping between a forest of overlapping absorption lines of Fe^+ and other species with low ionization potential in the rapidly expanding, cool, opaque shell. In particular, the sharp features below 2,000 Å do not correspond to emission lines expected from an ionized plasma, whereas the absorption minima are in general agreement with stronger Fe II multiplets. Recent calculations (S. Starrfield, personal communication, and ref. 11) that model the ultraviolet spectra of novae seem to confirm this general picture.

The fastest velocities observed in MWC560 $\sim -6,000 \text{ km s}^{-1}$ in the optical³ and $\sim -7,000 \text{ km s}^{-1}$ in the strong P Cyg absorptions⁵ associated with the ultraviolet h and k lines of Mg II are, as Margon¹² noted, 'far higher than those seen in symbiotic

stars, or even in the far more spectacular nova phenomenon'. P Cyg-type absorptions are attributed to optically-thick stellar winds. Indeed, the highest velocities observed in MWC560 with IUE in 1990 are twice as fast as those observed in visible light² in 1984. The unique spectra and velocities observed during the present outburst suggest that MWC560 may represent either a new type of cataclysmic variable star, a new type of symbiotic star, or a previously unobserved and possibly short-lived phase in the evolution of an interacting binary system. The emergence of a nova-like spectrum as observed on 1990 September 26 supports the hypothesis of a cataclysmic variable star; however, subsequent observations of MWC560 with IUE (1990 November 2 and 1990 December 18) showed that the ultraviolet spectrum continues to exhibit the characteristics of a cool, opaque shell. Also, the occurrence of a photometric outburst³ months before the emergence of the nova spectrum would apparently be unique in a nova. The long timescale for the evolution of the shell implied by these observations suggests that the symbiotic-star hypothesis may be more appropriate.

Typically, symbiotic stars show emission lines characteristic of a hot, rarefied nebula, such as forbidden lines of [O III], [O II] and [N II], as well as emission lines of C IV, N V, He II, C III] and the hydrogen Balmer series. In contrast, the pre-outburst observations^{1,2} of MWC560 showed Balmer lines, but not forbidden lines nor lines of high ionization state such as C IV, N V, or He II. Images of the symbiotic star R Aqr at optical and radio frequencies¹³ show extended bipolar nebulae which are attributed to ejection of large amounts of material on at least two occasions in past centuries. It is tempting to suppose that the recent appearance of a nova-like spectrum in MWC560 represents a similar episode of mass ejection; if so, imaging at high spatial resolution may reveal evidence of earlier shell-ejection episodes and, eventually, may resolve ejecta from the present event. If MWC560 represents a rapid evolutionary stage that precedes the conditions normally taken as characteristic of a symbiotic star, then the present event may be prelude to the

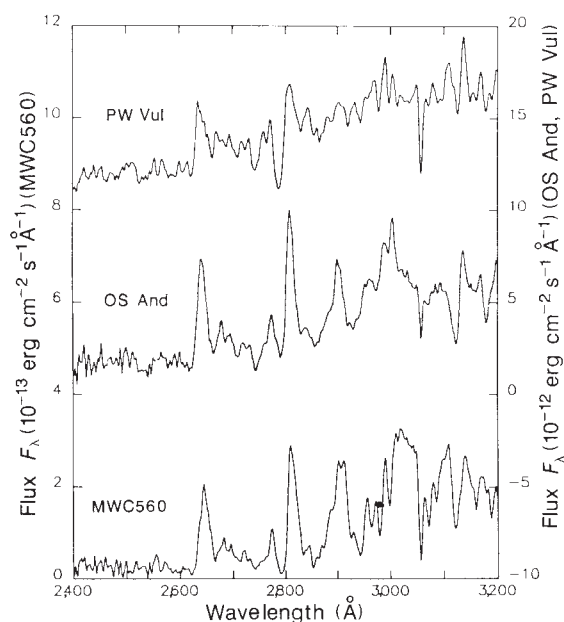


FIG. 3 Low-dispersion ultraviolet spectra (range 2,400–3,200 Å) of MWC560 (bottom) on 1990 September 26, nova OS And (centre) and nova PW Vul (top) obtained with the IUE long-wavelength spectrograph. Note the strong resemblances. Ordinates are as in Fig. 2; for purposes of presentation, the PW Vul flux has been multiplied by 4 and then increased by $1 \times 10^{-11} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ Å}^{-1}$.

emergence of a nebular spectrum like those of R Aqr and other fully-fledged members of the class of symbiotic stars. □

Received 5 November 1990; accepted 6 February 1991

- Merrill, P. W. & Burwell, C. G. *Astrophys. J.* **98**, 153–184 (1943).
- Bond, H. E., Pier, J., Pilachowski, C., Solvak, M. & Szkody, P. *Bull. Am. Astr. Soc.* **16**, 516 (1984).
- Tomov, T. *et al. Nature* **346**, 637–638 (1990).
- Sanduleak, N. & Stephenson, C. B. *Astrophys. J.* **185**, 899–913 (1973).
- Michalitsianos, A. G. *et al. Astrophys. J.* (in the press).
- Bode, M. F. & Evans, A. in *Classical Novae* (eds Bode, M. F. & Evans, A.) 163–186 (Wiley, New York, 1989).
- Stryker, L. L. *et al. in A Decade of UV Astronomy with the IUE Satellite* Vol. 1 (ed. Rolfe, E. J.) 149–152 (European Space Agency, Paris, 1988).
- McNaught, R. H. *et al. IAU Circ. No.* 3969 (1984).
- Kosai, H., Honda, M. & Wakuda, W. & Norimoto, Y. *IAU Circ. No.* 4281 (1986).
- Kriszunas, K. *et al. IAU Circ. No.* 4282 (1986).
- Wehrse, R., Hauschildt, P. H., Shaviv, G. & Starrfield, S. in *Evolution in Astrophysics* (ed. Rolfe, E. J.) 461–466 (European Space Agency, Paris, 1990).
- Margon, B. *Nature* **346**, 611 (1990).
- Michalitsianos, A. G. & Kafatos, M. in *The Symbiotic Phenomenon* (eds Mikolajewska, J., Friedjung, M., Kenyon, S. J. & Viotti, R.) 235–243 (Kluwer, Dordrecht, 1988).

ACKNOWLEDGEMENTS. We thank Y. Kondo for time on the IUE satellite and the IUE observatory staff for their support. S.P.M. thanks C. B. Stephenson and J. W. Truran for helpful discussions and acknowledges a letter from the late N. Sanduleak on the spectral classification of MWC560.

New measurement of the rate coefficient for the reaction of OH with methane

Ghanshyam. L. Vaghjiani* & A. R. Ravishankara†

Aeronomy Laboratory, National Oceanic and Atmospheric Administration, 325 Broadway, Boulder, Colorado 80303 and Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado 80309, USA

METHANE is an important greenhouse gas, whose concentration in the troposphere is steadily increasing. To estimate the flux of methane into the atmosphere and its atmospheric lifetime, its rate of removal needs to be accurately determined. The main loss process for atmospheric methane is the reaction with the hydroxyl radical OH. We have measured the rate coefficient for this reaction in carefully controlled experiments and found it to be smaller than currently accepted values. Our results indicate a longer CH₄ lifetime (by ~25%) and a correspondingly smaller flux (by ~100 Tg CH₄ yr⁻¹) than previously calculated.

Of all the trace tropospheric species (that is, excluding H₂O and CO₂) methane contributes most to the infrared heating of the atmosphere^{1,2}. Methane is also the most abundant hydrocarbon in the troposphere where it modulates the concentration of the OH free radical and serves as a source of CO. Transport of methane to the stratosphere provides a termination step, via the Cl+CH₄ reaction, for the chlorine-catalysed destruction of ozone. The oxidation of methane in the stratosphere is an important source of water vapour in this region. During the past decade^{3–5} the abundance of methane in the troposphere has been increasing at a rate between 16 and 13 parts per 10⁹ volume (p.p.b.v.) per year. The total input and the identities and strengths of the different atmospheric methane sources are not clearly defined. To understand the atmospheric effects of methane, and possibly to regulate it, we need these parameters. At present, the total flux of methane into the atmosphere is estimated from the measured steady-state abundance and the known removal rate of methane⁶. It has been generally accepted that the only process by which methane is chemically degraded

in the troposphere is the reaction with OH. Therefore, the rate coefficient, k_1 , for the reaction



is important in estimating the total flux of methane. The other loss processes, which are expected to be minor pathways, are surface deposition and reaction with Cl atoms in the lower stratosphere and upper troposphere.

A close examination of the available data^{7–10} shows that only in three investigations^{8–10} was k_1 measured below 298 K, the temperature region most important to the atmosphere. Only Davis *et al.*⁸ measured k_1 down to 240 K. Reaction 1 is slow. Therefore, at low temperatures, the presence of reactive impurities and occurrence of secondary reactions in laboratory systems can result in an overestimate of k_1 . We studied reaction 1 using an experimental method in which secondary chemistry could be minimized and the systematic errors reduced.

The pulsed photolysis-laser induced fluorescence (PP-LIF) method,¹¹ which is widely used for measuring free radical reaction rate coefficients, was employed here. The specific apparatus employed for the present study, the procedures for measuring OH reaction rate coefficients and the data analysis scheme are described elsewhere¹². The PP-LIF method relies on measuring the relative rate of change with time of the OH concentration, following its instantaneous production, in the presence of a known excess of methane. The bimolecular rate coefficient, k_1 , is obtained by measuring the first-order rate coefficient for the loss of OH, k' , in various concentrations of methane. The main

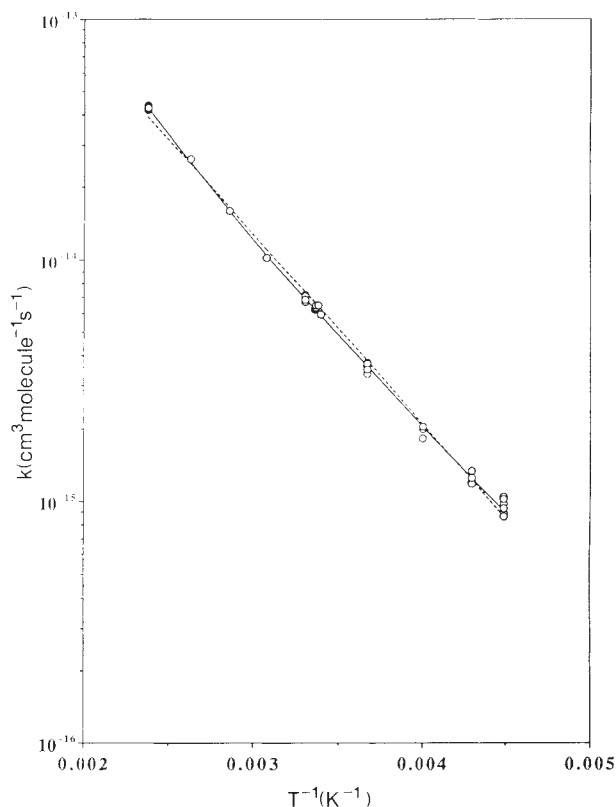


FIG. 1 Plot of k_1 (on a logarithmic scale) against T^{-1} . The 2σ precision of each point is less than its width on the graph. The data adequately fit the expression $k_1 = (2.94 \pm 0.34) \times 10^{-12} \exp[-(1815 \pm 30)/T] \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$ (equation 2, dashed line). Errors are 2σ and refer to precision only. The plot is slightly curved and a better fit is given by the expression $k_1 = 1.59 \times 10^{-20} T^{2.84} \exp(-978/T) \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$ (equation 3, solid line). We believe from the IR analysis of our sample that the curvature is not due to impurities. Such curvature has been seen in many other systems when k_1 has been measured over a large temperature range.

* Present address: University of Dayton Research Institute, Air Force Astronautics Laboratory, AL/LSCC, Edwards AFB, California 93523, USA.

† Address correspondence to this author at NOAA. Also affiliated with: Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado, USA.

TABLE 1 Experimental conditions and measured value of k_1

Temperature (K)	[OH] ₀ (10 ¹⁰ cm ⁻³)	[Photolyte]*	Range of [CH ₄] (10 ¹⁶ cm ⁻³)	k_1 (10 ⁻¹⁵ cm ³ molecule ⁻¹ s ⁻¹)§	Temperature (K)	[OH] ₀ (10 ¹⁰ cm ⁻³)	[Photolyte]*	Range of [CH ₄] (10 ¹⁶ cm ⁻³)	k_1 (10 ⁻¹⁵ cm ³ molecule ⁻¹ s ⁻¹)§
420	3.3	1.1	0.6–6.6	43.04 ± 0.47	273	15.3	5.1	10.3–60.1	3.64 ± 0.05
420	1.1	1.1	0.7–4.3	41.84 ± 0.61	273	188.0	15.7‡	9.9–41.3	3.67 ± 0.04
420	3.4	3.4	0.7–4.1	43.66 ± 0.56	273	100.0	9.0‡	10.2–67.0	3.75 ± 0.05
420	10.1	3.4	0.8–6.6	43.36 ± 0.63	273	45.0	19.4, 1.7†	13.8–55.2	3.71 ± 0.16
420	3.3	1.1	1.5–8.9	42.44 ± 0.57	273	45.0	47.1, 1.7†	20.4–79.7	3.38 ± 0.07
420	3.6	1.2	0.9–8.0	42.53 ± 0.52	273	14.1	4.7	13.1–33.9	3.52 ± 0.07
420	3.6	1.2	1.9–8.2	42.87 ± 0.52	273	15.3	5.1	5.5–32.5	3.50 ± 0.05
380	3.3	1.1	0.6–5.6	26.17 ± 0.42	250	15.3	5.1	14.2–98.9	1.98 ± 0.03
350	3.3	1.1	1.1–7.7	15.98 ± 0.11	250	14.7	4.9	6.0–45.7	2.04 ± 0.05
325	3.3	1.1	1.1–5.4	10.26 ± 0.15	250	14.4	4.8	8.7–28.0	1.82 ± 0.02
303	15.0	5.0	10.4–32.1	7.12 ± 0.38	233	9.9	3.3	14.3–70.1	1.21 ± 0.02
303	44.1	14.7	8.0–38.4	7.15 ± 0.11	233	3.3	1.1	13.7–73.2	1.19 ± 0.01
303	15.0	15.1	7.8–39.4	6.74 ± 0.08	233	1.1	1.1	7.8–36.4	1.18 ± 0.03
303	5.1	5.1	7.3–34.4	6.74 ± 0.13	233	10.2	3.4	10.1–63.3	1.18 ± 0.02
303	15.3	5.1	13.3–58.2	7.13 ± 0.18	233	9.6	3.2	9.9–61.1	1.25 ± 0.03
303	90.0	3.0†	18.7–74.1	6.84 ± 0.25	233	9.9	3.3	12.6–72.9	1.33 ± 0.02
					233	47.0	4.2‡	10.8–72.8	1.25 ± 0.02
298	15.3	5.1	15.1–71.5	6.20 ± 0.18	223	3.0	1.0	8.5–48.3	0.86 ± 0.02
298	15.0	5.1	7.2–48.9	6.30 ± 0.07	223	25.0	2.5‡	9.8–49.8	1.03 ± 0.02
					223	3.0	1.0	9.1–30.0	0.88 ± 0.04
					223	0.9	0.3	5.4–17.1	0.89 ± 0.04
297	28.0	2.3	15.6–44.2	6.36 ± 0.22	223	5.1	1.0	4.9–26.5	0.86 ± 0.04
297	29.0	2.4	15.1–46.9	6.47 ± 0.06	223	14.8	6.5‡	6.9–40.7	0.89 ± 0.03
					223	14.6	6.2‡	8.4–53.3	0.86 ± 0.02
295	4.8	4.8‡	17.4–44.6	5.92 ± 0.18	223	15.9	3.1‡	9.6–63.5	0.97 ± 0.03
					223	16.3	6.8‡	12.5–63.2	1.04 ± 0.02
					223	14.9	4.9‡	7.6–38.5	1.02 ± 0.02
					223	3.0	1.0	7.0–29.2	0.93 ± 0.03

* Unless otherwise marked, photolyte is H₂O and units are 10¹⁵ cm⁻³.

† Photolyte O₃, units 10¹³ cm⁻³. The O(¹D) was the source of OH. Concentration of OH is approximate. H₂O concentration was varied between 0 and 5) × 10¹⁶ cm⁻³.

‡ Photolyte H₂O₂, units 10¹³ cm⁻³.

§ Errors are 2σ.

Production of OH used one of three methods: (1) Broad band photolysis of H₂O between 165 nm (quartz cut-off) and 185 nm (water cut-off). Flash lamp energy 0.8–2.5 J; [OH]₀ ≈ 1–20 × 10¹⁰ cm⁻³; [H₂O] ≈ 1–50 × 10¹⁵ cm⁻³. Photolysis at λ > 120 nm (MgF₂ cut-off) was occasionally employed to check for photolysis of CH₄. Different focal length lenses were used to vary the fluence. (2) 248 nm KrF laser photolysis of H₂O₂. This method was used mostly to create very large concentrations of OH. Laser energies ~2–50 mJ cm⁻² pulse⁻¹; [OH]₀ ≈ 1–20 × 10¹¹ cm⁻³; [H₂O₂] ≈ 2–16 × 10¹³ cm⁻³. (3) 248 nm KrF laser photolysis of O₃ followed by reaction of O(¹D) with water or CH₄. This was used to test alternative methods for production of OH where CH₃ radicals are produced. Laser energy ~2 mJ cm⁻² pulse⁻¹; [OH]₀ ≈ 5–10 × 10¹¹ cm⁻³; [O₃] ≈ 2 × 10¹³ cm⁻³. OH detection was by pulsed-laser-induced fluorescence in the A ← X system. The excitation wavelength was 281.1 nm (A²Σ, ν' = 1 ← X²Π, ν'' = 0). Fluorescence was monitored at 312.2 (A²Σ, ν' = 1 → X²Π, ν'' = 0) and 306.4 nm (A²Σ, ν' = 1 → X²Π, ν'' = 0). Detection sensitivity was ~1 × 10⁷ cm⁻³ for 100 laser shots in the absence of CH₄. The buffer gas was helium or nitrogen (100–300 torr). Gas flow rates were 2–8 cm s⁻¹ through the reactor. Essentially all reactants are replaced between consecutive photolysis pulses (0.1 s apart at the higher flow rate).

OH removal process is reaction 1; minor contributions come from diffusional loss from the detection zone and reaction with bath gas impurities. The first-order rate coefficient for the latter two processes, k_d , is determined by measuring the OH loss rate in the absence of methane. We determine the concentration of methane by measuring the mass flow rates of methane, the diluent gas and the photolyte/diluent mixture (using calibrated electronic mass flow meters) and by measuring the total pressure (using a capacitance manometer). We can thus determine methane concentrations, and hence k_1 , very accurately (2σ ≤ 5%).

Table 1 describes the experimental conditions employed to measure k_1 . Fifty measurements of k_1 , between 223 and 420 K, are listed in Table 1. Variation of [OH]₀, due to variations in photolysis fluence and the concentration of the photolyte, did not affect the measured values of k_1 . The variations in photolysis fluence, flow rate of the gases, pressure and the nature of the diluent are not explicitly shown in the table. These variations also did not alter the measured values of k_1 . This demonstrates the absence of significant errors due to secondary reactions in our measurements, a point which will be discussed further. Our sample of methane was better than 99.99% pure. The levels of reactive impurities, propene and ethene, were found from

infrared measurement to be less than 0.2 p.p.m.v. and 0.7 p.p.m.v. respectively. Thus they contributed less than 2% to the measured value of k_1 even at 223 K, the lowest temperature of our study. The measured values of k_1 , and the fit of the data, are shown in Fig. 1. In one set of experiments (not shown in Table 1) we used a discharge-flow laser magnetic resonance system at 296 K to check the results from the pulsed photolysis experiments. The measured value of k_1 was 6.5 × 10⁻¹⁵ cm³ molecule⁻¹ s⁻¹, in excellent agreement with the flash photolysis results.

Our values of k_1 are smaller than those obtained earlier. The difference is close to the estimated uncertainties of the previous measurements and may not be considered significant. However, because of its significance to atmospheric chemistry, this difference is important. We believe that the principal reason for the smaller values measured here is that we have minimized the secondary reactions, especially the reaction of OH with the CH₃ product. To assess their contributions to the measured values of k_1 , we calculated the [OH] temporal profiles expected from the reactions shown in Table 2 by numerical integration. This scheme takes into account all reactions that could affect the measured [OH] temporal profiles. We carried out calculations for the concentrations used by Jeong and Kaufman¹⁰ and Davis

*et al.*⁸ as well as those used here. These two studies were chosen because they are the main sources for the currently recommended values of k_1 at lower atmospheric temperatures, they represent two of the most commonly used methods for measuring radical-molecule reaction rate constants (including k_1) and they report the initial conditions used. The rate coefficients for the loss of CH_3 radicals listed in Table 2 are the best available data for our experimental conditions. They were changed as appropriate to correspond to the conditions of other simulations. The simulations show that the rate coefficients reported earlier^{8,10} are overestimated by at least 15% at 273 K because of the contribution of the $\text{OH} + \text{CH}_3$ reaction. In our experiments, on the other hand, the contributions were less than 2%. If the rate coefficient for the $\text{CH}_3 + \text{H}$ reaction is taken to be the measured value¹³, then the importance of $\text{CH}_3 + \text{OH}$ reaction (and the associated error) increases in the experiments of Davis *et al.* Similarly, a decrease in the CH_3 wall loss or self-reaction rate would increase the errors of Jeong and Kaufman's experiments. The overestimation depends on the value of the rate coefficient for the $\text{OH} + \text{CH}_3$ reaction, for which there is only one measurement¹⁴; if the rate coefficient is smaller than this, the overestimation in earlier papers will be reduced. Without knowing the exact experimental conditions used by previous investigators, it is difficult to estimate the contributions of secondary chemistry. This modelling exercise is just to show that secondary reactions involving CH_3 were significant in most previous measurements. The variation of the measured k_1 with experimental parameters is the best evidence for the secondary reactions. Figure 2 shows the data of Davis *et al.* and indicates

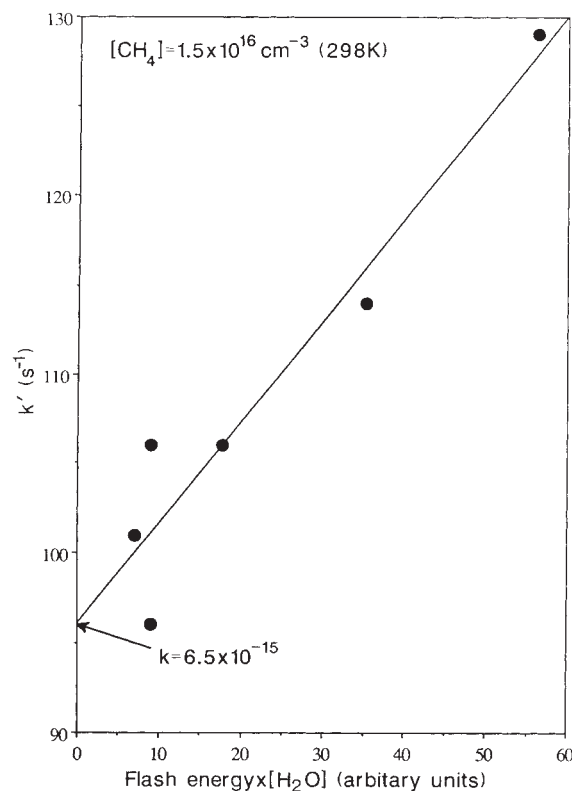


FIG. 2 Plot of the first order rate coefficient k' for the reaction of OH with CH_4 ($1.5 \times 10^{16} \text{ cm}^{-3}$) against the product of $[\text{H}_2\text{O}]$ and flash energy, from the experiments of Davis *et al.*⁸ (their table 1). The difference between the first order rate coefficients measured in the presence and absence of CH_4 is k' . The product of the flash energy and $[\text{H}_2\text{O}]$ is a relative measure of $[\text{OH}]_0$, the initial concentration of OH. To keep the photolysis wavelength the same, we have only plotted those results for which a sapphire window was used. The intercept, which is the value of k' in the absence of the $\text{OH} + \text{CH}_3$ secondary reaction, yields $k_1 = 6.5 \times 10^{-15} \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$, in good agreement with our value. The slope of the plot is related to the rate of the $\text{OH} + \text{CH}_3$ reaction, but cannot be evaluated because we do not know the absolute concentrations of OH that were used.

TABLE 2 Kinetic scheme and rate coefficients used in the Runge-Kutte integration of $[\text{OH}]$ temporal profiles at 273 K

Reaction	Rate coefficient*, $\text{cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$	Reference
$\text{CH}_4 + \text{OH} \rightarrow \text{CH}_3 + \text{H}_2\text{O}$	9.0×10^{-16}	
$\text{CH}_3 + \text{OH} \rightarrow \text{Products}^\dagger$	9.0×10^{-11}	14, 15
$\text{CH}_3 + \text{H} + \text{M} \rightarrow \text{CH}_4 + \text{M}$	2.0×10^{-10}	13
$\text{CH}_3 + \text{CH}_3 + \text{M} \rightarrow \text{C}_2\text{H}_6 + \text{M}$	5.5×10^{-11} to zero ‡	17
$\text{CH}_3 \rightarrow \text{loss}$	$20 (\text{s}^{-1})$	assumed §
$\text{H} \rightarrow \text{loss}$	$20 (\text{s}^{-1})$	assumed §
$\text{H} + \text{OH} + \text{M} \rightarrow \text{H}_2\text{O}$	4.0×10^{-12}	16, 17
$\text{OH} + \text{OH} + \text{M} \rightarrow \text{H}_2\text{O}_2$	3.1×10^{-12}	16, 17

* For simulations of experiments carried out in a flow tube¹⁰ at 3 torr, or in flash photolysis apparatus⁸ at ~ 25 torr, the pressure-dependent rate coefficients were changed appropriately. Many of the rate coefficients are set to minimize the effect of secondary reactions in these studies.

† We assume that the product of the $\text{CH}_3 + \text{OH}$ reaction does not react with OH. If it did, it would lead to further overestimation of k_1 .

‡ The rate coefficient for the $\text{CH}_3 + \text{CH}_3$ reaction depends on pressure and bath gas. If this reaction is fast, it suppresses the $\text{OH} + \text{CH}_3$ reaction. To simulate our experiments, this value was set to zero to maximize $[\text{CH}_3]$, whereas when we simulated the experiments of Jeong and Kaufman and Davis *et al.* it was set to the maximum value to minimize $[\text{CH}_3]$.

§ In the experiments involving the flash photolysis of H_2O , $[\text{H}]_0$ is assumed to be the same as $[\text{OH}]_0$.

that k_1 was indeed overestimated in their experiments. A similar plot of our data at 223 K (the temperature at which the effect would be most evident) yields an intercept which is within one standard deviation of the mean at that temperature. Furthermore, the correlation coefficient, R^2 , of our plot is only 0.47. Any contributions from the impurities in the samples used by Jeong and Kaufman¹⁰ and Davis *et al.*⁸ would only make k_1 larger.

The chief effect of revising the value of k_1 is to alter the CH_4 budget. All atmospheric processes that are now evaluated from the measured atmospheric CH_4 concentration, such as the current flux of CH_4 to the stratosphere, will be unaffected by the change in k_1 . Generally, k_1 values recommended by the NASA¹⁶ or the CODATA¹⁷ panels are used for atmospheric lifetime and budget calculations. A crude way to appraise the atmospheric significance of our results is to examine the values of k_1 at 273 K, the weighted mean temperature at which the removal of a species such as methane, with a lifetime of several years, will take place. The NASA and CODATA values, at 273 K are respectively 24% and 29% higher than those predicted by equation (3) in Fig. 1. Our results indicate that the lifetime of methane in the troposphere is $\sim 25\%$ longer and the flux $\sim 25\%$ ($\sim 100 \text{ Tg yr}^{-1}$) smaller than currently accepted values. This change in flux is comparable to or less than every one of the main sources^{6,18}. More detailed modelling calculations, which take into account the effects of feedback, use temperature and OH fields and include transport, can better assess the effect of this revision in k_1 . Such a calculation by Valentin and Crutzen (personal communication) shows that the methane lifetime is $\sim 20\%$ longer when our values are used instead of the CODATA recommendation. They also show that using our values 490 Tg yr^{-1} of CH_4 is removed through OH reaction, as opposed to 590 Tg yr^{-1} of CH_4 using the CODATA recommendation. Conversely, if the flux of methane into the troposphere is indeed 590 Tg yr^{-1} , then there is a missing sink for CH_4 in the troposphere. $\square$

Received 19 December 1990; accepted 13 February 1991.

1. Ramanathan, V., Cicerone, R. J., Singh, H. B. & Kiehl, J. T. *J. geophys. Res.* **90**, 5547–5566 (1985).
2. Mitchell, J. F. B. *Rev. Geophys.* **27**, 115–139 (1989).
3. Steele, L. P. *et al. J. Atmos. Chem.* **5**, 125–171 (1987).
4. Khalil, M. A. K. & Rasmussen, R. A. *Envir. Sci. Technol.* **24**, 549–553 (1990).
5. Blake, D. R. & Rowland, F. S. *Science* **239**, 1129–1131 (1988).
6. Cicerone, R. J. & Oremland, R. S. *Global biogeochem. Cycles* **2**, 299–327 (1988).
7. Ravishankara, A. R. *Ann. Rev. Phys. Chem.* **39**, 367–394 (1988).

8. Davis, D. D., Fischer, S. & Schiff, R. *J. chem. Phys.* **61**, 2213-2219 (1974).
9. Margitan, J. J., Kaufman, F. & Anderson, J. G. *Geophys. Res. Lett.* **1**, 80-81 (1974).
10. Jeong, K. M. & Kaufman, F. *J. phys. Chem.* **86**, 1808-1815 (1982).
11. Hancock, G. *Measurements of Thermal Rate Data in Modern Gas Kinetics: Theory, Experiment and Applications* (eds Pilling, M. J. & Smith, I. W. M., Blackwell, Oxford, 1987).
12. Vaghjani, G. L. & Ravishankara, A. R. *J. phys. Chem.* **93**, 1948-1959 (1989).
13. Brouard, M., McPherson, M. T. & Pilling, M. J. *J. phys. Chem.* **93**, 4047-4059 (1989).
14. Sworski, T. J. & Hochanadel, C. J. & Ogren, P. J. *J. phys. Chem.* **84**, 129-134 (1980).
15. Tsang, W. & Hampson, R. F. *J. phys. chem. Ref. Data* **15**, 1087-1279 (1986).
16. DeMore, W. B. *et al.* *NASA JPL Publ.* 90-1 1990.
17. Atkinson, R. *et al.* *J. phys. chem. Ref. Data* **18**, 881-1097 (1989).
18. Taylor, J. A., Brasseur, G., Zimmerman, P. & Cicerone, R. J. *J. geophys. Res.* **96**, 3013-3044 (1991).

ACKNOWLEDGEMENTS. We thank K. Valentin and P. Crutzen for helpful discussions and calculations; R. J. Cicerone for valuable discussions and the pre-print of his paper; D. Bopegedera and J. B. Burkholder for analysis of our methane samples; T. Gierczak and R. Talukdar for the flow tube measurements; and C. J. Howard and D. L. Albritton for reading the manuscript. This research was carried out under the Radiatively Important Trace Species component of the NOAA Climate and Global Change Program.

Effects of ice coverage and ice-rafted material on sedimentation in the Fram Strait

Dieter Hebbeln & Gerold Wefer

Universität Bremen, Geowissenschaften, Postfach 330440,
2800 Bremen 33, FRG

AS little is known about pelagic sedimentation processes in Arctic environments¹, the interpretation of biological and chemical processes, as well as the reconstruction of ancient conditions, including those in the glacial North Atlantic, is difficult. Here we provide sediment-trap results, which show that the position of the sea-ice boundary significantly influences the particle flux. The seasonal variability of the particle flux differed markedly in the various sediment-trap sites in Fram Strait, depending on the behaviour of the sea ice. Under complete ice cover, sedimentation is very low, whereas maximum sedimentation is found at the ice margin. The highest particle flux observed, showing a large lithogenic component, was observed at the ice edge where the water was warmer ($>2^{\circ}\text{C}$). We find that high biogenic opal fluxes are characteristic of the summer ice margin, indicating that the sedimentary record of opal fluxes may allow the position of ice margins in the past to be reconstructed.

In the Fram Strait, the narrow gateway between the North Atlantic and the Arctic Ocean, areas with constant and seasonal

ice coverage, as well as ice-free areas are present (Fig. 1). The large variability in seasonal ice coverage is caused by a complex system of warm (West Spitsbergen) and cold (East Greenland) currents (Fig. 1). As there is such a large spatial and seasonal variability in ice cover, the Fram Strait is particularly well suited for the study of the relation between ice coverage and particle sedimentation.

Three one-year trap deployments were positioned east-west across the Fram Strait (Fig. 1) site FS4 was permanently ice-covered, whereas another (FS3) was ice-covered for about six months. The third site (SP2) was located at the winter ice margin, close to the Spitsbergen shelf edge. The 20-cup-collector sediment traps, with a collection area of 0.5 m^2 , were attached to moored arrays at depths ranging from 1,110 to 1,488 m. Depending on the expected ice cover, samples were taken at intervals of between 11.5 and 23 days. Field work and laboratory analyses are described elsewhere². The flux rates of individual components are given in Fig. 2.

Site FS4 was located in the East Greenland Current, which is characterized by year-round water temperature $<0^{\circ}\text{C}$. The annual particle flux is 2.9 g m^{-2} (Fig. 2a), and was the lowest of the three sites, with a maximum from August to October and a minimum between January and April. Site FS3 was in the Central Fram Strait in an area where the ice margin fluctuates for 70% of the year³. During this time, the water temperature is near 0°C . The annual particle flux was 60.5 g m^{-2} (Fig. 2b), with the daily flux ranging from a maximum of $700\text{ mg m}^{-2}\text{ d}^{-1}$ in July, to between 11 and $21\text{ mg m}^{-2}\text{ d}^{-1}$ for December and January. The eastern part of the Fram Strait is ice-free most of the year, but during 1989, pack ice reached site SP2 in January. Because the water temperature of the West Spitsbergen Current is higher than 0°C , the ice melts rather quickly. The total annual flux at this site was 149.9 g m^{-2} (Fig. 2c), and was the highest measured at the three sites. The seasonal variation in the flux differed from those observed further west in the time at which the flux increased. Between July and December, a more or less continuous flux of 80 to $210\text{ mg m}^{-2}\text{ d}^{-1}$ was measured. In January the total flux then increased, reaching $1,147\text{ mg m}^{-2}\text{ d}^{-1}$, followed by a decrease reaching $\sim 500\text{ mg m}^{-2}\text{ d}^{-1}$ in May. A change in the flux composition was also observed; from July to December the particle flux was dominated by fecal pellets, whereas from January to April quartz grains of silt- to fine sand-size were common.

The most important factors influencing sedimentation are therefore the presence of ice and the surface water temperatures.

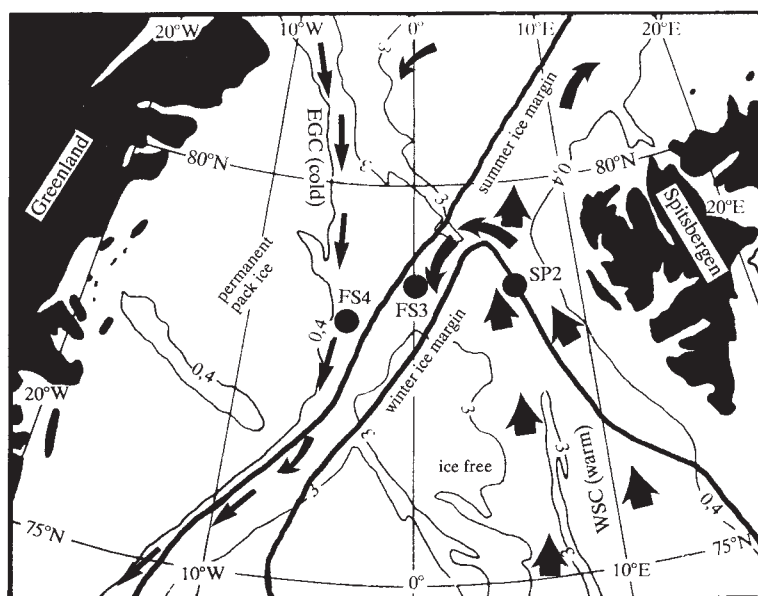


FIG. 1 Positions of trap sites FS4 ($78^{\circ}26.5'N$, $04^{\circ}05.7'W$, water depth 1,763 m, trap depth 1,191 m), FS3 ($78^{\circ}45.8'N$, $00^{\circ}10.7'E$, water depth 2,487 m, trap depth 1,488 m) and SP2 ($78^{\circ}53.4'N$, $06^{\circ}44.5'E$, water depth 1,661 m, trap depth 1,110 m) in the Fram Strait. The average position of ice margins^{1,3} and the cold East Greenland Current (EGC) and the warm West Spitsbergen Current (WSC) are also shown.

There is little ice melting in cold waters ($<0^{\circ}\text{C}$) and consequently only small amounts of sediment were collected at site FS4, even though the ice contains large amounts of terrigenous material⁴. In contrast, at site SP2, in the warm waters (always $>2^{\circ}\text{C}$) of the West Spitsbergen Current, large amounts of ice-rafted material were collected from melting sea ice. In the central Fram Strait (site FS3), particle sedimentation was determined

by the complex hydrographic situation in the contact zone between the major current systems. Particle flux was low during the winter because of the extensive ice coverage and lack of biological productivity. During summer, when the ice margin was positioned above the trap, melting ice released terrigenous material into the water. Wind-forced vertical mixing along the ice edge enhanced biological productivity⁵, which also caused an increase in the particle flux. Because average surface-water temperatures at the ice margin in the central Fram Strait are $\sim 0^{\circ}\text{C}$, the flux rates of ice-rafted material at this site cannot reach those characteristic of the warmer waters at site SP2. Although the data presented here covers only one year, the relationship between the results obtained and the environmental conditions suggests that the data set is representative. These seasonal flux patterns were also observed in the central Fram Strait⁶ and off West Spitsbergen during other years (D.H. and G.W., unpublished data).

The maxima in particle sedimentation at sites FS4 and FS3 are related to the plankton development in the surface waters. At site FS4, the extent of open water is greatest in early autumn. Light availability is then highest and enhanced primary production can take place. At site FS3 in the Central Fram Strait, the total flux was much higher compared to site FS4 because of higher biological productivity and the release of ice-rafted material. The higher productivity can probably be attributed to higher light availability and an increased nutrient supply which may result from ice-edge-upwelling⁵. The time delay between the flux maxima at sites FS4 and FS3 (September as opposed to July–August, Fig. 2a, b) is caused by the time lag between the position of the ice margin in the area of site FS3 (moving back and forth during summer), and the maximum extent of open water areas at site FS4 (largest during early autumn).

A different sedimentation pattern is observed at site SP2 (Fig. 2c). The relatively uniform flux between July and December is supplied by the northward flowing West Spitsbergen Current. The strong increase in particle sedimentation in January coincides with the arrival of the ice margin over the trap site. We assume that during this season, the ice was melted quickly by the warm West Spitsbergen Current ($>2^{\circ}\text{C}$), thereby releasing large amounts of ice-rafted material into the water and producing a sudden increase in particle sedimentation. We estimate that $>50\%$ of the total annual flux at this site was released by rapidly melting ice between January and April. The qualitative change in the flux composition mentioned earlier supports this contention. It is possible that a substantial amount of material was resuspended during winter storms on the Spitsbergen shelf, as suggested by the high percentage of benthic foraminifera in these samples (J. Carstens, personal communication). Another possible source of material could be the West Spitsbergen Current. Results of a sediment-trap deployment off Bear Island at 75°N and 11°E show an excess lithogenic sedimentation of 6.6 g m^{-2} during winter, which is probably due to the outflow of dense winter water from the Barents shelf⁷. For the same period, the excess lithogenic sedimentation for SP2 is $\sim 75\text{ g m}^{-2}$, indicating that such a winter water outflow does not explain the high flux rates observed at SP2 during winter. From observations of sea ice an input of 75 g m^{-2} from melting ice floes seems probable^{4,8}. The retreat of the ice and the return of daylight increases primary production, as indicated by the increase in the flux of organic carbon and opal. Large particle fluxes can therefore persist, even though there is no supply of ice-rafted material.

From the trap data, the sedimentation regime in the Fram Strait can be characterized by five different flux situations: (1) closed ice cover above the East Greenland Current; (2) ice cover with open water areas above the East Greenland Current; (3) ice-margin conditions; (4) open water above the West Spitsbergen Current; and (5) ice cover above the West Spitsbergen Current. Figure 3 shows the average daily particle sedimentation for these five situations.

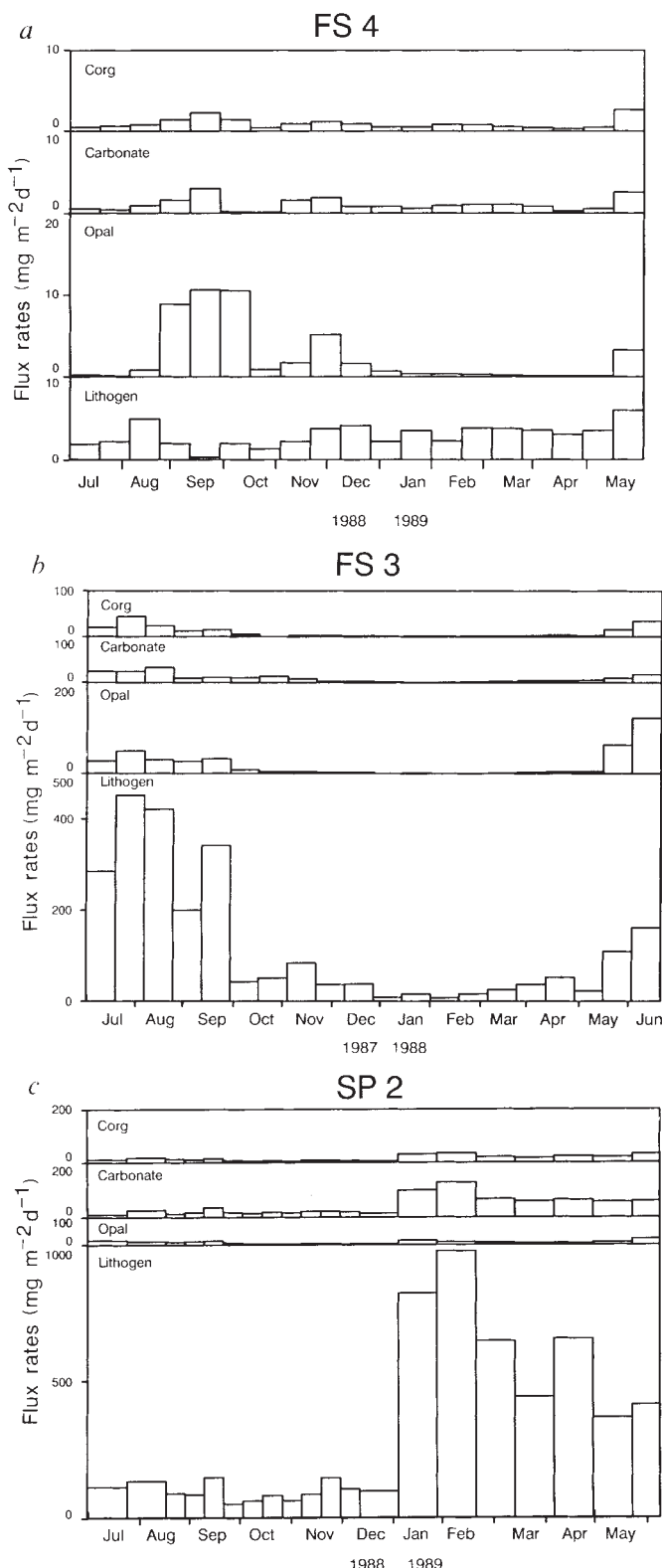
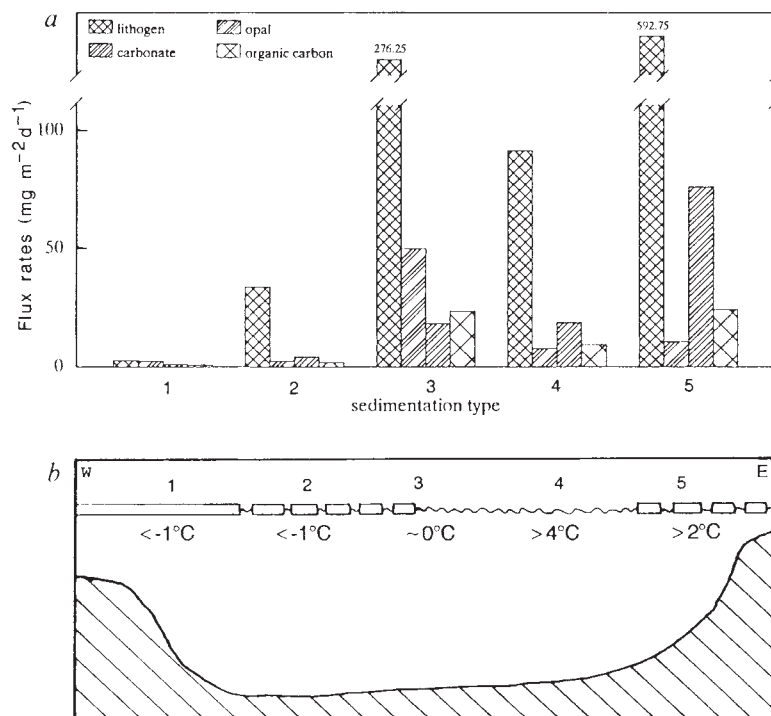


FIG. 2 Flux rates of the size fraction $<1\text{ mm}$ at the three sites in the Fram Strait; a, site FS4 in the western part; b, site FS3 in the central part; and c, site SP2 in the eastern part of Fram Strait (note different scales).

FIG. 3 *a*, Average sedimentation in the five sedimentation regimes of the Fram Strait: (1) closed ice cover above the East Greenland Current; (2) ice cover with open water areas above the East Greenland Current; (3) ice-margin conditions, (4) open water above the West Spitsbergen Current; and (5) ice cover above the West Spitsbergen Current. *b*, Profile through Fram Strait showing different sedimentation situations.



Situations (1) and (2) in the East Greenland Current are characterized by very low fluxes, which seems typical for areas of permanent ice coverage, because similar results were obtained from the ice-covered Weddell Sea⁹. Because no melting can occur in the cold water (<-1°C) the high portion of lithogenic material in (2) is probably released from the ice by wave wash-off or turning of ice floes⁴. The ice-margin situation (3) is marked by the highest observed opal flux, which can be attributed to favourable conditions for biological productivity. For this as well as for the ice-covered West Spitsbergen Current (situation (4)), the high lithogenic flux results from the release of ice-rafted material by melting. That is also the reason for the high carbonate flux in (5), which is partly composed of ice-transported old carbonates. A high opal flux due to enhanced primary production cannot occur because of the lack of daylight in (5). As indicated by the dominance of fecal pellets in (4) above the West Spitsbergen Current, the high lithogenic flux is caused by high abundances of zooplankton and a high availability of lithogenic material, probably originating from the Barents Sea and transported by the West Spitsbergen Current.

Ice transport will also be found around polynas. Owing to the mode of origin of a polyna, the resulting flux pattern will be similar to that found in (2) in the case of a wind-driven coastal polyna (for example, the East Greenland Shelf polyna¹⁰) where water temperatures remain cold, or to type (3), if a polyna is caused by the upwelling of water warm enough to melt the ice (for example, the Weddell Sea polyna¹¹).

Our sediment-trap data show that high biogenic opal fluxes may be related to summer-ice-margin conditions (Fig. 3). That enhanced opal production that can be recorded may be seen in sediment cores from the Skagerrak/Norwegian Sea and Fram Strait, where diatom maxima were related to past positions of the polar front¹², which during summer roughly follows the ice margin. Opal maxima in sediment cores should therefore allow the average position of the ice margin in summer in the past to be reconstructed. □

- Berner, H. & Wefer, G. in *Geological History of the Polar Oceans: Arctic versus Antarctic* (eds Bleil, U. & Thiede, J.) 317-335 (Kluwer, Dordrecht, 1990).
- Pfirman, S., Lange, M.A., Wollenburg, I. & Schlosser, P. in *Geological History of the Polar Oceans: Arctic versus Antarctic* (eds Bleil, U. & Thiede, J.) 187-211 (Kluwer, Dordrecht, 1990).
- Smith, W. O. Jr., Baumann, M. E. M., Wilson, D. L. & Aletsee, L. *J. geophys. Res.* **92**, 6777-6786 (1987).
- Honjo, S., Manganini, S. J., Karowe, A. & Woodward, B. L. *Tech. Rep. WHOI-87-17* (Woods Hole Oceanographic Institution, Woods Hole, 1987).
- Honjo, S., Manganini, S. J. & Wefer, G. *Deep-Sea Res.* **35**, 1223-1234 (1988).
- Wollenburg, I. & Pfirman, S. *Berichte FB Geowissenschaften Univ. Bremen* **6**, 52-57 (1989).
- Fischer, G. et al. *Nature* **335**, 426-428 (1988).
- Wadhams, P. in *The Nordic Seas* (ed. Hurdle, B. G.) 21-78 (Springer, New York, 1986).
- Comiso, J. C. & Gordon, A. L. *J. geophys. Res.* **92**, 2819-2833 (1987).
- Stabell, B. *Geol. Rundsch* **75**, 175-184 (1986).
- Vinje, T. E. *Norsk Polarinst. Arbok* 1975 163-174 (1977).

ACKNOWLEDGEMENTS. We thank J. Meincke, J. Carstens and the crew of RV *Polarstern* for their help in deploying and recovering the sediment traps, and M. Botros and W. H. Berger for critical reading of the manuscript and helpful comments. This research was supported by the Bundesminister für Forschung und Technologie (Projekt Framstraße).

Mixing-layer distortion at the confluence of channels of different depth

James L. Best* & André G. Roy†

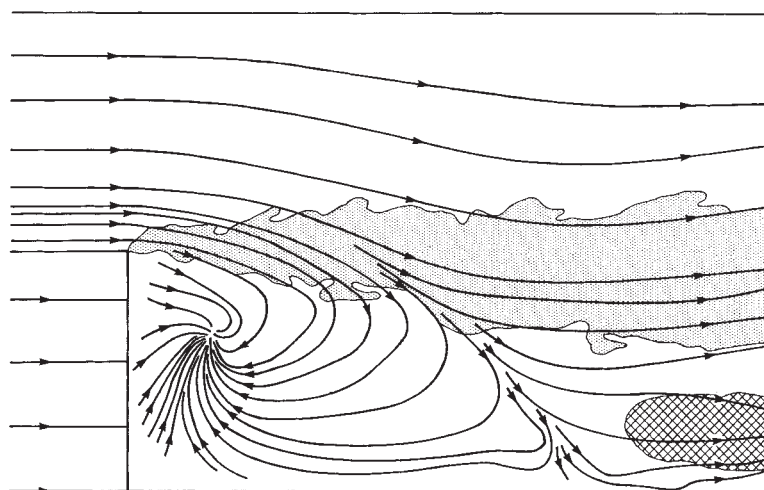
* Department of Earth Sciences, University of Leeds, Leeds LS2 9JT, UK
† Département de Géographie, Université de Montréal, CP 6128, Montréal, Québec H3C 3J7, Canada

CHANNEL confluences have received considerable recent attention in the fields of geomorphology¹⁻³, sedimentology⁴⁻⁹ and hydraulic engineering¹⁰⁻¹³. Where channels join, rapid changes in fluid velocity, turbulence intensity, channel hydraulic geometry and bed geometry may occur¹⁴⁻¹⁷. Previous models of junction dynamics have tended to assume that the confluent channels are of equal depth^{10,18}, a condition that may be found only rarely in natural junctions; more often, the depths of confluent channels are different¹⁹. This discordance is commonly the result of the formation, at the mouth of each confluent channel, of bars that possess steep avalanche faces which dip into a central junction scour^{1,2,5,20}. Here we suggest that when channels of different depth join, the mixing layer that forms at their confluence is distorted by

Received 18 October; accepted 21 January 1991.

- Honjo, S. in *Polar Oceanography* (ed. Smith, W.O.) 687-739 (Academic, New York, 1990).
- Wefer, G. et al. in *Geological History of the Polar Oceans: Arctic versus Antarctic* (eds Bleil, U. & Thiede, J.) 363-379 (Kluwer, Dordrecht, 1990).

FIG. 1 Bed streamlines and spreading of the mixing layer at a Reynolds number of 12,670. Bed streamlines show the extent of the separation zone behind the step and the entrainment of fluid from the deeper channel. Stippled shading indicates the maximum extent of the mixing layer, and cross-hatching denotes the zone of fluid upwelling produced by deformation of the mixing layer.



interacting with a separation zone which forms in the lee of the mouth of the shallower channel. This mixing-layer distortion imparts a complex three-dimensional flow to the junction, promoting vertical fluid upwelling and significantly enhancing rapid mixing within the flow of the shallower tributary, while retarding the rate of mixing across the flow of the deeper channel. Suspended sediments or contaminants within the deeper channel may be transferred rapidly across the flow of the shallower tributary, upwelling downstream of the junction.

To investigate the dynamics of laterally mixing flows with different depths, we have conducted a series of experiments in a flume consisting of two parallel channels, each 0.135 m wide and 3 m long, which merged at the downstream end of a blunt-ended splitter plate. Flow in one channel was made shallower by inserting a gently tapered ramp, 1.1 m in length, which terminated in a 0.05-m high 90° negative step at the end of the splitter plate (Fig. 2 insert). Flow depths were kept constant, being 0.10 and 0.05 m in the deeper and shallower channels respectively, and were allowed to mix freely downstream from the splitter plate in a channel 0.27 m wide and 3 m long. The results presented here are based on visualization of the mixing flows by injecting dye both at the negative step and on the water surface; bed streamlines were constructed from traces of water-based paint placed on the flume floor. Velocity measurements were made using a DANTEC two-component laser Doppler anemometer. Measurements were made at three mean Reynolds numbers, and in each experiment there was a slight difference in velocity between the two confluent flows, the shallower channel being between 0.83 and 0.99 times the velocity in the deeper channel.

The mixing of two flows of different depths essentially involves interaction between the mixing layer that forms where the flows meet and a zone of separated flow formed in the lee of the step. Bed streamlines and the time-averaged position of the mixing layer at a mean Reynolds number of 12,670 (Fig. 1) illustrate that the separation zone possesses a complex three-dimensional recirculation as a result of interaction with the mixing layer. A nodal point of upwelling fluid is present within the separation zone approximately one step height downstream of the step edge. The length of the separation zone, defined by its maximum downstream extent along the wall of the flume, remained constant at 3.5 step heights for all three Reynolds numbers studied. Deformation of the mixing layer closely follows the curvature of bed streamlines around the separation zone (Fig. 1). At each Reynolds number, the rate of mixing-layer growth decreases at a certain distance downstream (Fig. 2). The distance to this point of change in rate decreases at higher Reynolds number, reflecting the increasing entrainment of fluid into the separation zone from the mixing layer and deeper channel. Another possible

consequence of increased fluid entrainment into the separation zone is that, at a given position downstream, the mixing layer becomes narrower at higher Reynolds numbers (Fig. 2). Therefore, as the Reynolds number increases, suspended sediments or contaminants will increasingly tend to become entrained into the separation zone, enhancing mixing across the flow of the shallower channel but retarding mixing across the flow of the deeper channel downstream of the confluence.

Fluid within the mixing layer is also considerably distorted in three dimensions through interaction with the separation zone. The mixing layer is characterized by Kelvin-Helmholtz instabilities which may pair, split and coalesce as they convect downstream²¹. In the presence of the separation zone, these vortices are gradually distorted as they travel downstream, their bases being deformed inwards towards the separation zone as revealed by the bed streamlines (Fig. 1). Vortices that originate with vertical axes in the initiation zone of the mixing layer are gradually deformed and adopt a vorticity that has its axis in the cross-stream plane (Fig. 3). The vortex bases are drawn progressively towards the wall of the shallower channel and, as vortex stretching and break-up occurs, fluid from the mixing layer (and the deeper channel) upwells within the body of the shallower-channel flow downstream from the separation zone (Figs 1 and

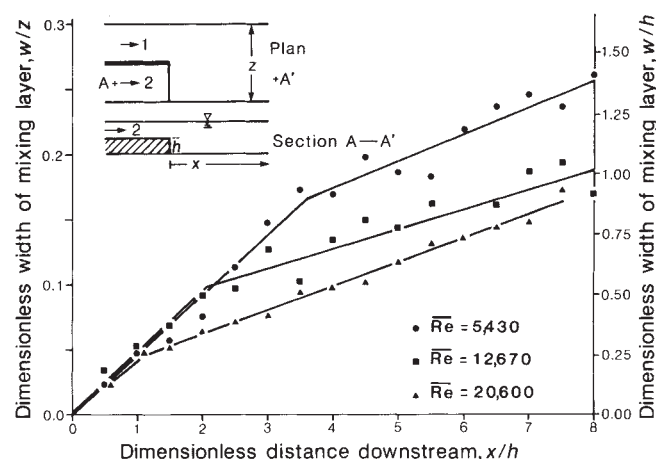


FIG. 2 Spreading rate of the mixing layer for the three mean Reynolds numbers examined. The width of the mixing layer is made dimensionless by dividing either by the channel width z or by the step height h . The downstream distance, x , has its origin at the step edge and is made dimensionless by dividing by the step height.

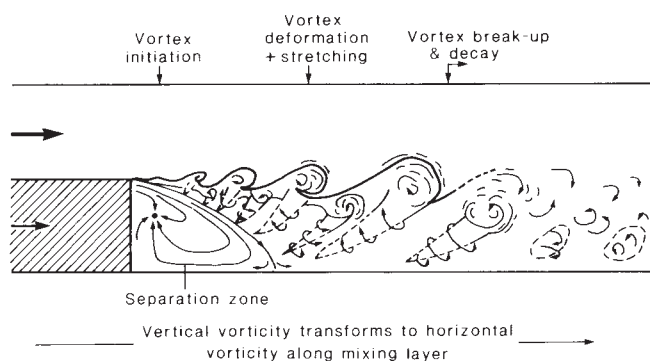


FIG. 3 A schematic model of mixing-layer distortion at the confluence of two flows of different depth. This planform diagram illustrates that deformation of the vertical mixing-layer vortices leads to the production of horizontal vorticity, enhanced mixing and fluid upwelling downstream of the confluence.

3). In Fig. 3 we illustrate a schematic model of this process of vortex and mixing-layer distortion.

Such complex three-dimensional deformation of the mixing layer is characteristic of all the unequal-depth mixing flows that we have studied. Although we have examined only zero degree confluence angles, mixing-layer distortion such as that described here can be expected at any junction in which the confluent channels are of different depths. The orientation of the separation zone is controlled by the planform geometry of the negative step, which in natural channels is formed by avalanche faces of tributary mouth bars which dip into the central confluence scour^{1,5,6}. Consequently, the degree of entrainment of deeper-channel fluid into the separation zone and the position of the zone of upwelling will be dependent on the position of the tributary mouth bars, this itself being controlled by the junction angle and the ratio of discharges between the confluent channels^{2,5,18}. This process of mixing-layer distortion and fluid upwelling may be invoked to explain patterns of fluid mixing at natural river junctions^{22,23} where fluid from one channel is observed to upwell within the body of the other tributary flow downstream from the confluence.

The process we have outlined here has many implications. The dispersal of sediment at junctions across channel towards the shallower tributary may be greatly enhanced by such a depth differential between the confluent channels. This may influence erosion, through the position of the mixing layer and its associated higher Reynolds stresses, and deposition through control of the location of both zones of upwelling and the principal sediment transport pathways. Helical flow cells, which have been described at river channel confluences^{1,20}, may originate from such depth-differential-controlled processes and not from any inherent helical flow where flows of equal depth combine. Attempts at modelling the near-field of side discharges in effluent dispersal²⁴ must also account for this mixing-layer distortion, which may significantly increase the potential for mixing across the shallower channel at a junction while simultaneously decreasing the rate of mixing across the deeper channel. Thus, effluents suspended in a deep channel may be rapidly transferred across the flow of a shallower channel when the two flows merge. □

Received 19 November 1990; accepted 18 February 1991.

- Mosley, M. P. *J. Geol.* **84**, 535–562 (1976).
- Ashmore, P. E. & Parker, G. *Water Resour. Res.* **19**, 342–402 (1983).
- Roy, A. G. & Bergeron, N. *Geomorphology* **3**, 99–112 (1990).
- Klassen, G. J. & Vermeer, K. *Proc. Int. Conf. Fluvial Hydraul.* Budapest, 1–14 (1988).
- Best, J. L. *Sedimentology* **35**, 481–498 (1988).
- Mosley, M. P. & Schumm, S. A. *Econ. Geol.* **72**, 691–694 (1977).
- Reid, I., Best, J. L. & Frostick, L. E. in *Floods: Hydrological, Sedimentological and Geomorphological Implications* (eds Beven, K. & Carling, P.) 135–150 (Wiley, Chichester, 1989).
- Bryant, I. D., Holyoak, D. T. & Mosely, K. A. *Proc. Geol. Ass.* **94**(4), 321–343 (1983).
- Best, J. L. & Brayshaw, A. C. *J. geol. Soc. Lond.* **142**, 747–755 (1985).

- Soong, H. K. thesis, Univ. Connecticut (1976).
- Hager, W. H. *J. Hydraul. Engng* **115**, 243–259 (1989).
- Sutherland, A. J. in *Proc. 9th Aus. Fluid Mech. Conf.* 259–263 (Univ. of Auckland, 1986).
- Chu, V. H. & Babarutsi, S. *J. Hydraul. Engng* **114**, 1257–1274 (1988).
- Richards, K. S. *Water Resour. Res.* **16**, 241–244 (1980).
- Best, J. L. & Reid, I. *J. Hydraul. Engng* **110**, 1588–1594 (1984).
- Roy, A. G. & Woldenberg, M. J. *J. Geol.* **94**, 401–411 (1986).
- Roy, A. G. & Roy, R. *Earth Surf. Proc. Landf.* **13**, 77–84 (1988).
- Best, J. L. *Soc. Econ. Palaeontologists and Mineralogists spec. Publ.* **39**, 27–35 (1987).
- Kennedy, B. A. *Earth Surf. Proc. Landf.* **9**, 153–173 (1984).
- Ashmore, P. E. *Earth Surf. Proc. Landf.* **7**, 201–225 (1982).
- Winant, C. D. & Browand, F. K. *J. Fluid Mech.* **63**, 237–255 (1974).
- Sternberg, H. O. *The Amazon River of Brazil* 74 (Steiner, Wiesbaden, 1975).
- Roy, A. G. & De Serres, B. *Bull. Soc. Géog. Liège* **25**, 113–127 (1989).
- McGuirk, J. J. & Rodi, W. *J. Fluid Mech.* **86**, 761–781 (1978).

ACKNOWLEDGEMENTS. We thank the Earth Sciences workshop at Leeds for construction of the mixing-flow apparatus. Travel for A.G.R. was made possible through a grant from FCAR; this paper was prepared while J.L.B. held a Royal Society/NSERC Anglo-Canadian Scientific Exchange Scheme Fellowship at the Université de Montréal that was also supported by the Foundation for Canadian Studies. We are grateful for the thoughtful comments of Jim Pizzuto.

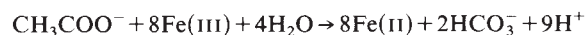
Microbial reduction of uranium

Derek R. Lovley, Elizabeth J. P. Phillips, Yuri A. Gorby & Edward R. Landa

Water Resources Division, 430 National Center, US Geological Survey, Reston, Virginia 22092, USA

REDUCTION of the soluble, oxidized form of uranium, U(VI), to insoluble U(IV) is an important mechanism for the immobilization of uranium in aquatic sediments and for the formation of some uranium ores^{1–10}. U(VI) reduction has generally been regarded as an abiological reaction in which sulphide, molecular hydrogen or organic compounds function as the reductant^{1,2,5,11}. Microbial involvement in U(VI) reduction has been considered to be limited to indirect effects, such as microbial metabolism providing the reduced compounds for abiological U(VI) reduction and microbial cell walls providing a surface to stimulate abiological U(VI) reduction^{1,12,13}. We report here, however, that dissimilatory Fe(III)-reducing microorganisms can obtain energy for growth by electron transport to U(VI). This novel form of microbial metabolism can be much faster than commonly cited abiological mechanisms for U(VI) reduction. Not only do these findings expand the known potential terminal electron acceptors for microbial energy transduction, they offer a likely explanation for the deposition of uranium in aquatic sediments and aquifers, and suggest a method for biological remediation of environments contaminated with uranium.

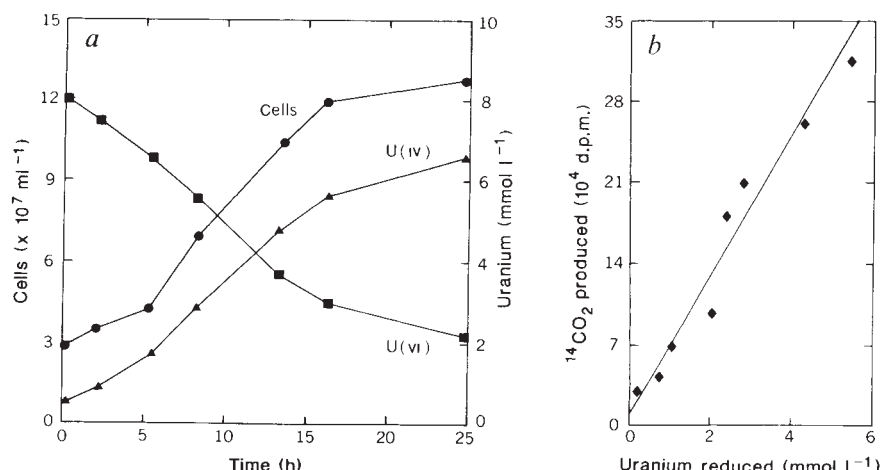
It is becoming increasingly clear that the reduction of metals in anaerobic environments is often the result of the direct enzymatic reduction by bacteria^{14,15}. For example, the Fe(III)-reducing microorganism, strain GS-15, grows under anaerobic conditions by enzymatically coupling the oxidation of acetate to carbon dioxide with the reduction of Fe(III) to Fe(II)^{16,17} according to:



Thermodynamic calculations have indicated that, per electron transferred, acetate oxidation coupled to U(VI) reduction has the potential to yield more than twice the energy that is available from Fe(III) reduction⁸. When GS-15, grown on acetate and Fe(III), was inoculated into an anaerobic medium with acetate as the sole electron donor and U(VI) as the potential electron acceptor, U(VI) was reduced to U(IV) over time (Fig. 1a). Growth coincided with U(VI) reduction and stopped as U(VI) became depleted. When [2-¹⁴C]-acetate was incorporated into the medium, ¹⁴CO₂ was generated in direct proportion to U(VI) reduction (Fig. 1b). In a separate experiment in which acetate concentrations were measured in cultures growing with U(VI) reduction, U(VI) and acetate loss over time were linearly related (correlation coefficient $r=0.9$) with a U(VI)/acetate ratio of

FIG. 1 Cell numbers, and U(vi) and U(iv) concentrations over time (a), and relationship between $^{14}\text{CO}_2$ production from $[2-^{14}\text{C}]$ -acetate and U(vi) reduction (b) when GS-15 was inoculated into a medium with acetate as the electron donor and U(vi) as the electron acceptor. GS-15 obtained energy to support growth by reducing U(vi) to U(iv) while oxidizing acetate to carbon dioxide.

METHODS. Cells grown anaerobically at 30 °C in an acetate-Fe(III)-citrate medium¹⁷ were inoculated into a similar medium but with Fe(III) replaced with 10 mmol l⁻¹ uranyl chloride. Acetate concentrations were: 10 mmol l⁻¹ (a) and 3.5 mmol l⁻¹ (including 0.5 µCi of $[2-^{14}\text{C}]$ -acetate (53 mCi/mmol)) (b). Cell numbers¹⁷ and $^{14}\text{CO}_2$ produced²⁹ were determined as previously described. All manipulations for the determination of U(vi) and U(iv) concentrations were carried out in an anaerobic chamber. Subsamples were acidified with 12 N HCl to provide a final concentration of 4 N HCl. In (a), 2 ml of each acidified subsample were immediately added to a glass column (10-cm length, 1-cm inner diameter) containing Dowex AG1X8. As previously determined⁸, on this column U(vi) is retained in 4 N HCl whereas U(iv) is not. The column was washed with 20 ml of 4 N HCl to collect the U(iv) and then washed with 20 ml of 0.1 N HCl to elute the U(vi). The concentrations of uranium in the U(vi) and U(iv) fractions were determined by directly aspirating the samples into a directly coupled plasma

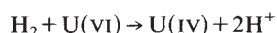


3.6:1. Given that a small proportion of the acetate metabolized would be incorporated into cells, these results indicate that GS-15 can obtain energy for growth by oxidizing acetate with the reduction of U(vi) to U(iv) according to:



Energy conservation from acetate oxidation coupled to U(vi) reduction should be through electron-transport and oxidative phosphorylation, because there is no known mechanism to generate net ATP through substrate level phosphorylation with acetate as the substrate. Subsequent studies have indicated that, in GS-15, electron transport to U(vi) proceeds through a respiratory chain containing cytochrome *b*₅₅₈ (Y.G., manuscript in preparation). GS-15 is considered to reduce U(vi) directly rather than indirectly through reduction of Fe(III) with subsequent reduction of U(vi) by Fe(II), because Fe(II) did not reduce U(vi) in the growth medium and U(vi) was rapidly reduced in cell suspensions that had been repeatedly washed and resuspended in iron-free buffer.

One other organism, *Alteromonas putrefaciens*, is known to obtain energy for growth from electron transport to Fe(III)^{18,19}. *A. putrefaciens* grew with H₂ as the sole electron donor and U(vi) as the electron acceptor (Fig. 2). In parallel studies, the stoichiometry of H₂ consumption and U(vi) reduction was consistent with the reaction:



Other electron donors for Fe(III) reduction by *A. putrefaciens* (formate, lactate, pyruvate) also supported U(vi) reduction.

Neither H₂ nor any of the organic electron donors for U(vi) reduction by GS-15 or *A. putrefaciens* reduced U(vi) in the absence of the organisms (Fig. 3 and data not shown). Furthermore, reduction of U(vi) in washed cell suspensions of GS-15 or *A. putrefaciens* was much faster than the reduction of U(vi) in the presence of a high concentration (1 mmol l⁻¹) of sulphide (Fig. 3). With extended incubation, cell suspensions of GS-15 reduced U(vi) concentrations below 0.4 µmol l⁻¹.

These results demonstrate that, although abiological reduction of U(vi) by sulphide, H₂ or organic compounds is typically considered to be the mechanism for U(vi) reduction in sedimentary environments^{1,2,5,11}, enzymatic reduction of U(vi) by microorganisms using U(vi) as a terminal electron acceptor is also

possible. Furthermore, microbial reduction of U(vi) has the potential to proceed much more rapidly than abiological U(vi) reduction. Previous evidence for the potential of microorganisms to reduce U(vi) was the report that, in the presence of U(vi), crude cell-free extracts of *Micrococcus lactilyticus* consumed H₂ in quantities that were consistent with U(vi) reduction to U(iv)²⁰. But these extracts reduced a wide range of metals with no evidence that the metal reduction was an enzymatic reaction or had any physiological significance in whole cells¹⁵.

To evaluate more fully the relative potential for microbiological and abiological U(vi) reduction in reduced sediments, U(vi) was added to highly reduced methane-producing, sulphide-containing sediments of the Potomac River (Fig. 4). U(vi) reduction was much more rapid and extensive in the biologically

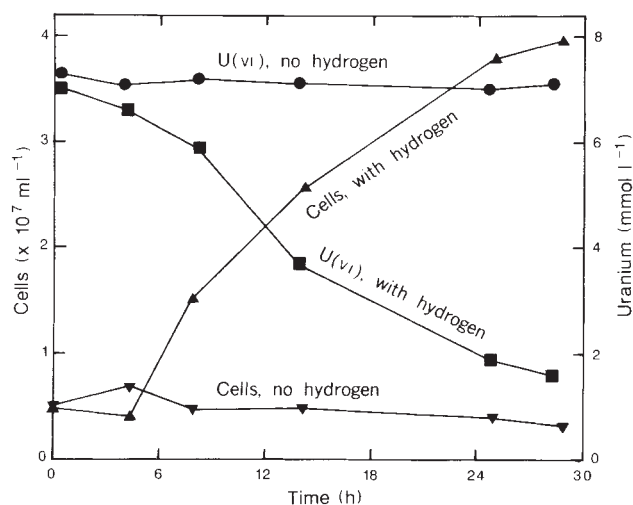
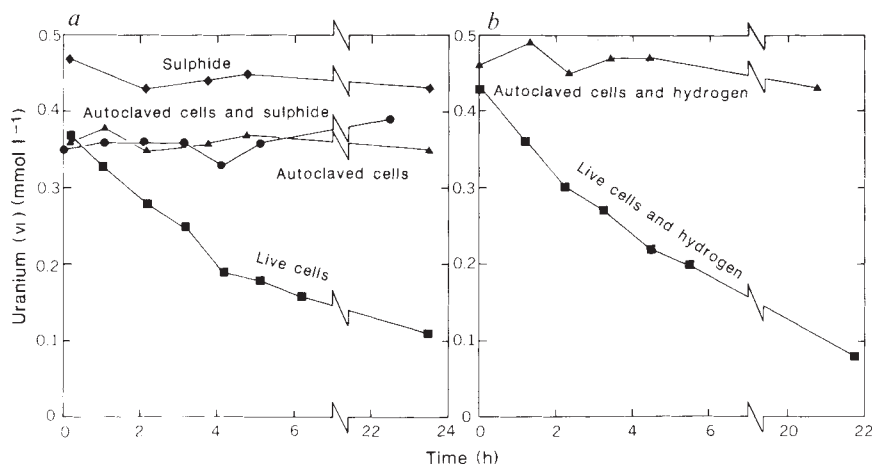


FIG. 2 Cell numbers and U(vi) and U(iv) concentrations over time when *Alteromonas putrefaciens* was inoculated into a U(vi)-containing medium with or without H₂. *A. putrefaciens* grew with the reduction of U(vi) to U(iv) in the presence of H₂, but not with H₂ omitted.

METHODS. Cells grown anaerobically at 30 °C with H₂ (60 kPa) as the electron donor and Fe(III)-citrate as the electron acceptor¹⁸ were inoculated into a similar medium but with Fe(III) replaced with 10 mmol l⁻¹ uranyl acetate. Controls had no added H₂. U(vi) was determined as in Fig. 1b.

FIG. 3 U(VI) concentrations over time in the presence of various combinations of potential U(VI) reductants and cells of GS-15 (a) or *Alteromonas putrefaciens* (b). U(VI) was only reduced in the presence of actively metabolizing cells.

METHODS. Washed cells of Fe(III)-grown GS-15 (a) or *A. putrefaciens* (b) were suspended at 30 °C under N₂-CO₂ (80:20) in 10 ml of bicarbonate buffer (NaHCO₃, 0.025g; pH 6.7) containing uranyl acetate (0.4 mmol l⁻¹) to provide ~100 ng of cell protein per ml. Sodium sulphide (1 mmol l⁻¹) or H₂ (56 kPa) were added as noted. Aliquots were removed over time, flushed with N₂-CO₂ to remove sulphide if necessary, and U(VI) was determined as in Fig. 1b.



active sediments than in sediments in which the microorganisms were inactivated with heat. The microorganisms responsible for the U(VI) reduction were not identified, but these sediments do contain Fe(III)-reducing microorganisms which rapidly become active when an electron acceptor is added²¹. These results further suggest that models for U(IV) deposition in sedimentary environments should consider the possibility that microbial populations that have grown up with Fe(III), or possibly other electron acceptors, can reduce U(VI) entering these environments.

Geochemical evidence from a number of environments is consistent with U(VI) reduction by Fe(III)-reducing microorganisms leading to U(IV) deposition. For example, in marine sedi-

ments, deposition of U(IV) typically takes place within the sediment zone in which microorganisms oxidize organic matter with the reduction of Fe(III)^{4-8,10}. Sulphide is unlikely to be the U(VI) reductant in these instances as sulphide does not accumulate in the Fe(III)-reducing zone of marine sediments. Furthermore, sulphide does not appear to reduce U(VI) in marine waters^{9,22,23}. Abiological reduction of U(VI) by organic matter or H₂ in the Fe(III)-reducing zone of sediments also seems unlikely because organic-matter reduction of U(VI) is restricted to high temperature (>120 °C)^{24,25} and, as shown above, at low temperatures, H₂ does not readily reduce U(VI) in the absence of microbial activity.

The common observation of U(IV) accumulations in the bleached 'reduction spots' of otherwise red, Fe(III)-rich rocks^{26,27} provides another example in which uranium deposition is likely to be associated with the activity of Fe(III)-reducing bacteria. The lack of red colour in the reduction spots is due to localized reduction of Fe(III) which, recent evidence suggests, is the result of microbial Fe(III) reduction^{26,28}. Even for environments such as sandstone- or roll-type uranium deposits, in which there is a co-accumulation of sulphide and U(IV) minerals, there is no direct evidence for sulphide reduction of U(VI) and other reduction mechanisms cannot be ruled out¹².

The potential for uranium contamination of surface and ground waters through uranium mining activities, irrigation of agricultural lands and disposal of nuclear wastes is an environmental concern. The results presented here suggest that, in many instances, it may be possible to immobilize uranium contamination by stimulating microbial U(VI) reduction in aquatic sediments or ground water. GS-15 and other Fe(III)-reducing microorganisms can oxidize important organic contaminants with Fe(III) as the electron acceptor^{29,30}. Thus, in the case of 'mixed wastes', which contain both organic contaminants and radioactive metals, the activity of U(VI)-reducing bacteria might be able to couple the decomposition of toxic organic compounds with the immobilization of uranium. Other radioactive metals, such as plutonium and technetium, which have multiple redox states and are insoluble in the reduced form, could potentially be treated in a similar manner. □

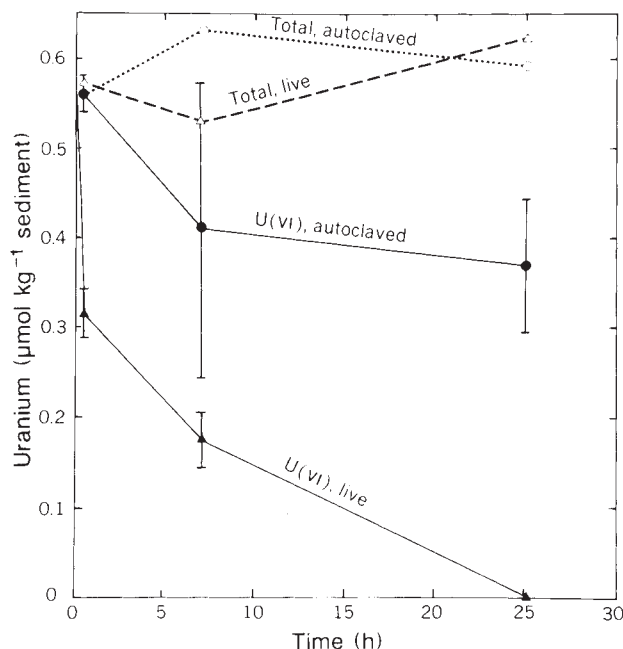


FIG. 4 Concentrations of U(VI) and total uranium over time in live and sterilized sediments. U(VI) was reduced much faster and more completely in live sediments, which reduced U(VI) below the detection limit of 2 nmol l⁻¹ within 25 h.

METHODS. Uranyl acetate was added to anaerobically incubated, methane-producing sediments from the Potomac River in which sulphate had already been reduced to sulphide²¹. Sediments were autoclaved four times (121 °C, 1 h) for abiotic controls. Over time, subsamples (~0.3 g) were removed and extracted under anaerobic conditions in 5 ml of 100 mmol l⁻¹ bicarbonate for 30 min. U(VI) was measured as in Fig. 1b. To determine total uranium in the extracts, U(IV) was first oxidized to U(VI) by bubbling the extract in air for 5 min.

Received 8 November 1990; accepted 15 January 1991.

- Jensen, M. L. *Econ. Geol.* **53**, 598-616 (1958).
- Hostettler, P. B. & Garrels, R. M. *Econ. Geol.* **57**, 137-167 (1962).
- Turekian, K. K. & Bertine, K. K. *Nature* **229**, 250-251 (1971).
- Bonatti, E., Fisher, D. E., Joensuu, O. & Rydell, H. S. *Geochim. cosmochim. Acta* **35**, 189-201 (1971).
- Langmuir, D. *Geochim. cosmochim. Acta* **42**, 547-569 (1978).
- Kadko, D. *Earth planet. Sci. Lett.* **51**, 115-131 (1980).
- Colley, S. & Thomson, J. *Geochim. cosmochim. Acta* **49**, 2339-2348 (1985).
- Cochran, J. K., Carey, A. E., Sholkovitz, E. R. & Surprenant, L. D. *Geochim. cosmochim. Acta* **50**, 663-680 (1986).
- Anderson, R. F., LeHuray, A. P., Fleisher, M. Q. & Murray, J. W. *Geochim. cosmochim. Acta* **53**, 2205-2213 (1989).
- Wallace, H. E. *et al. Geochim. cosmochim. Acta* **52**, 1557-1569 (1988).
- Maynard, J. B. *Geochemistry of Sedimentary Ore Deposits* (Springer, New York, 1983).

12. Taylor, G. H. in *Biogeochemical Cycling of Mineral-Forming Elements* (eds Trudinger, P. A. & Swaine, D. J.) 485–514 (Elsevier, New York, 1979).
13. Mohagheghi, A., Updegraff, D. M. & Goldhaber, M. B. *Geomicrobiol. J.* **4**, 153–173 (1985).
14. Ehrlich, H. L. *Geomicrobiology* (Marcel Dekker, New York, 1990).
15. Ghiorse, W. C. in *Biology of Anaerobic Microorganisms* (ed. Zehnder, A. J. B.) 305–331 (Wiley, New York, 1988).
16. Lovley, D. R., Stolz, J. F., Nord, G. L. & Phillips, E. J. P. *Nature* **330**, 252–254 (1987).
17. Lovley, D. R. & Phillips, E. J. P. *Appl. environ. Microbiol.* **54**, 1471–1480 (1988).
18. Lovley, D. R., Phillips, E. J. P. & Lonergan, D. J. *Appl. environ. Microbiol.* **55**, 700–706 (1989).
19. Myers, C. R. & Nealson, K. H. *Science* **240**, 1319–1321 (1988).
20. Woolfolk, C. A. & Whiteley, H. R. *J. Bacteriol.* **84**, 647–658 (1962).
21. Lovley, D. R. & Phillips, E. J. P. *Appl. environ. Microbiol.* **53**, 2636–2641 (1987).
22. Anderson, R. F., Fleisher, M. Q. & LeHuray, A. P. *Geochim. cosmochim. Acta* **53**, 2215–2224 (1989).
23. Anderson, R. F. *Uranium* **3**, 145–164 (1987).
24. Breger, I. A. in *Formation of Uranium Ore Deposits* 99–124 (International Atomic Energy Agency, Vienna, 1974).
25. Nakashima, S., Disnar, J. R., Perruchot, A. & Trichet, J. *Geochim. cosmochim. Acta* **48**, 2321–2329 (1984).
26. Hoffman, B. A. *Chem. Geol.* **81**, 55–81 (1990).
27. Harrison, R. K. *G.B. geol. Surv. Bull.* **52**, 1–26 (1975).
28. Lovley, D. R., Chapelle, F. H. & Phillips, E. J. P. *Geology* **18**, 954–957 (1990).
29. Lovley, D. R. et al. *Nature* **339**, 297–299 (1989).
30. Lovley, D. R. & Lonergan, D. J. *Appl. environ. Microbiol.* **56**, 1858–1864 (1990).

ACKNOWLEDGEMENTS. We thank Ruth Smith and Ray Van Hoven for technical assistance and Ron Oremland, Tom Kraemer, Marty Goldhaber, Rich Reynolds and Richard Smith for helpful discussions. This study was supported by the US Geological Survey Toxic Waste/Groundwater Contamination Programme.

Evidence for volcanic eruption on the southern Juan de Fuca ridge between 1981 and 1987

William W. Chadwick Jr*, Robert W. Embley† & Christopher G. Fox†

* OSU/CIMRS, Hatfield Marine Science Center, Newport, Oregon 97365, USA

† PMEL/NOAA, Hatfield Marine Science Center, Newport, Oregon 97365, USA

THE formation of new ocean crust at mid-ocean ridges is known to be a discontinuous process in both space and time, but little is known about the frequency and duration of eruptions along an active ridge segment. Here we present evidence, from Sea Beam surveys and underwater photography, for the eruption of lavas along a segment of the Juan de Fuca ridge between 1981 and 1987. Although previous studies have inferred volcanic activity on ridges in areas where recent seismicity or young lava flows have been observed^{1–4}, none has yet had direct evidence to date such a recent submarine eruption. The temporal coincidence between this eruptive episode and the megaplumes (huge, sudden emissions of hot mineral-laden water) observed over this part of the ridge^{5,6} in 1986 and 1987 supports previous suggestions^{5–10} that megaplumes are caused by sea-floor spreading events.

The lavas that were erupted during the 1980s occur as a series of pillow mounds and ridges between 45°00.5' N and 45°09.5' N along the northern Cleft segment (Fig. 1). The evidence for the recent eruption of these lavas was first discovered through a discrepancy between microbathymetry derived from a towed camera system and bathymetric charts based on Sea Beam surveys. Three camera tows (collected in 1989 from the National Oceanic Atmospheric Administration (NOAA) ship *Discoverer*) over the southernmost pillow mound (mound 1, Fig. 1) showed glassy lava flows forming a 35-m-high hill, ~1 km in diameter. Earlier Sea Beam bathymetry (based on data collected in May 1981 from the NOAA ship *Surveyor*) shows no such hill, but instead a gentle slope to the east (Fig. 2a). The Sea Beam and camera-tow bathymetry correspond closely in other areas.

Other surveys confirm the recent appearance of mound 1. Bathymetry from Sea Beam surveys conducted over the same area by the research vessel *Atlantis II* in September 1987 and by the *Discoverer* in August 1990 shows a 25-m-high mound in the exact location of the glassy lavas delineated by the camera

tows (Fig. 2b). Mound 1 was also imaged by a SeaMARC I sidescan survey in August 1987 from the *Discoverer*, and appears as an unfractured positive relief feature which covers pre-existing fractures (Fig. 3). The date of eruption of mound 1 is further restricted by a single Sea Beam swath collected during a camera tow by the *Surveyor* in June 1983. Although the data only sample the western edge of mound 1, the coverage is adequate to show that the 1983 swath has the same pre-eruption depth contours as the 1981 Sea Beam survey, and does not show the contours of the new lava mound. These Sea Beam surveys definitely constrain the time of the eruption of mound 1 to between 1981–1987, and probably to between 1983–1987.

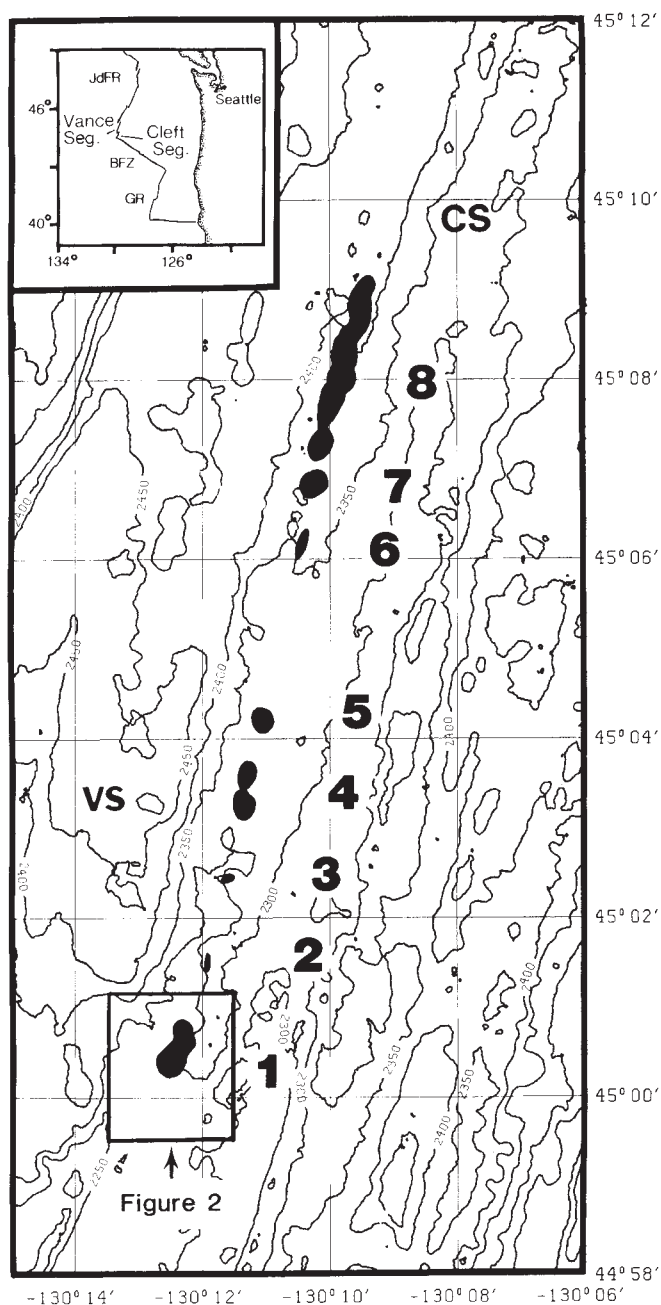


FIG. 1 Sea Beam bathymetric map of overlap region between Cleft and Vance segments, southern Juan de Fuca ridge (50-m contours); CS is northern end of Cleft segment, VS is southern end of Vance segment. Solid black areas show locations of new pillow lava mounds. Numbers to right of each mound are for identification in text. Box around mound 1 shows outline of Fig. 2. On inset map, GR is Gorda ridge, BFZ is Blanco Fracture Zone, and JdFR is Juan de Fuca ridge.

Further comparisons of the 1981, 1987 and 1990 Sea Beam data sets to the north of mound 1 reveal that several other mounds appear in the 1987/1990 data that were not present in 1981 (Fig. 1). No significant changes are evident between the 1987 and 1990 Sea Beam surveys, but the 1987 survey only covered mounds 1, 7 and 8, and not the intervening area, whereas the coverage of the 1990 survey was complete. Therefore the time of eruption of mounds 4 and 5 is only constrained to between 1981 and 1989 (the year of the camera tow). Mounds 2, 3 and 6 are glassy pillow mounds mapped by camera tows in 1989, but are apparently too small to be resolved by Sea Beam, so their eruption date can only be limited to before 1989. Most of the volume of new lava was erupted between 1981 and 1987; whether the features were constructed simultaneously or in discrete events within this time period cannot be determined from these data. The entire string of new eruptive sites lies roughly in the direction N 15° E over a length of 16 km. This line of recent volcanic activity is clearly along the north extension of the Cleft segment⁷, but it is also within the zone of overlap of the Cleft and Vance segments, and the new eruptions extend into the axial valley of the Vance segment (Fig. 1).

Significant changes in the sea floor between Sea Beam surveys can be inferred qualitatively from obvious changes in the shape of bathymetric contours. Figure 2a and b shows bathymetric contour maps made from 1981 and 1990 Sea Beam data, using identical mathematical gridding and automated contouring routines. Mound 1 can be distinguished as a new landform in Fig. 2b which is not present in Fig. 2a. We also used a quantitative technique to compare digital bathymetric grids from various Sea Beam data sets so that we could objectively identify areas of significant bathymetric changes. Figure 2c was produced by the following steps: (1) the 1990 Sea Beam grid was subtracted from the 1981 grid; (2) the differences were weighted in inverse proportion to the slope of the sea floor at each grid cell; (3) a final grid was created consisting of only those original (unweighted) depth differences for which the slope-weighted differences exceeded a preselected threshold. The differences must be weighted because a given error in location (due primarily to navigation error) will create a larger error in depth differences over a steep slope than over a gentle slope. For Fig. 2c, a threshold was chosen such that the outline of the area of significant differences best corresponded to the edge of the glassy lavas as determined from camera tows. As this threshold value is slope-weighted, it does not correspond to a single depth difference; the smallest depth differences above the threshold range from 5 to 15 m. We

believe the technique to be reliable because in every case where camera-tow coverage is available, all of the new mounds defined by the Sea Beam difference grids occur precisely where camera tows indicate fresh glassy lavas. Mound 8 is just beyond the limit of the 1989 camera tows and will be the target of future camera surveys.

Because of its finite acoustic footprint and limited depth resolution, Sea Beam cannot detect all new lava flows. A few of the smallest patches of fresh lavas mapped by the towed camera system in 1989 do not produce a detectable change in bottom topography between the 1981 and 1990 Sea Beam surveys. Nevertheless, these lavas are identical in morphology and sediment cover to the other new lavas, and all the mounds occur along a linear trend. Therefore, we suspect all the glassy lava mounds were erupted during the 1981 to 1987 period.

A total volume for the new lava mounds of 0.05 km³ is calculated from the gridded Sea Beam differences. For comparison, this is similar to the volume produced during the 1977 eruption of Kilauea volcano, Hawaii (0.04 km³), which lasted 18 days¹¹. During the 9-year rifting episode in northeast Iceland (1975–1984)^{12,13}, a total volume of 0.1–0.2 km³ of basalt was erupted in 9 out of 21 separate intrusion events; the volume of intruded dykes was much larger, ~1.0 km³.

Many apparently young, glassy lavas have been observed along the mid-ocean-ridge system, but it has not been possible to determine exact ages. Now we have a benchmark for the character of 'zero-age' lava. Bottom photographs taken in 1989 and observations from the submersible *Alvin* on dive 2262 in 1990 (Fig. 3) show that the new lavas are similar in appearance and sediment cover. Each new mound is primarily an accumulation of pillow lavas. The lava flows consist of high-standing bulbous and elongate pillows with corrugated surfaces, surrounded by cylindrical lobes with smooth glassy surfaces that resemble subaerial pahoehoe toes. The difference in microrelief of these two surface textures gives them a consistent difference in apparent sediment cover. The smooth pahoehoe toes have no sediment veneer, whereas small amounts of sediment are present along the extrusion marks on the bulbous pillows. Small sediment pockets between pillows are present in some areas. There is a sharp contrast in light reflectivity and sediment cover between the new lava flows and the surrounding older lavas.

In addition to the grey-brown pelagic sediment, there are two types of yellow deposits on the surface of the fresh lavas. One type consists of small globules of yellow material commonly located along small cooling cracks on individual pillows and

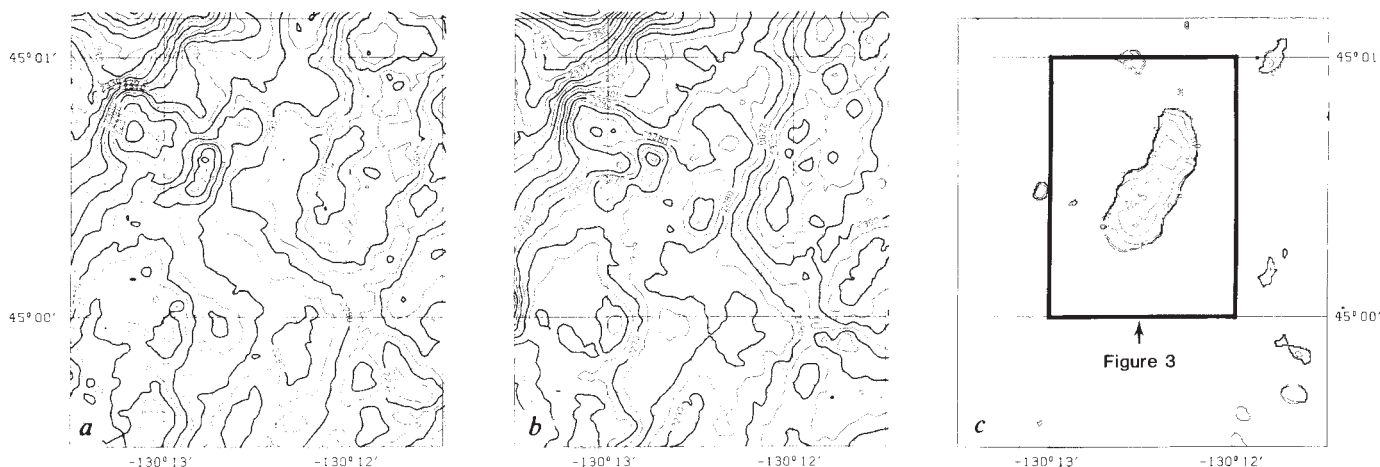


FIG. 2 Sea Beam bathymetric maps (5-m contours) of area including southern-most new lava mound (mound 1 in Fig. 1). a, Pre-eruption bathymetry from survey in 1981. b, Post-eruption bathymetry from survey in 1990. Mound 1 is a new feature near centre of map. c, Map of the significant

depth differences between 1981 and 1990 Sea Beam surveys (see text for discussion), showing outline and shape of mound 1. This corresponds exactly with the location of a mound of fresh, glassy pillow lavas based on camera tows, *Alvin* dives and sidescan sonar data (Fig. 3).

ubiquitous throughout the mounds. This material seems to form from the direct interaction of sea water and hot lava. A review of the literature and data from other areas indicates that this type of deposit is not generally seen on older pillows and may be a good indication of very young submarine lava flows. The second type of deposit appears as a fine dusting of yellow-tinged material, and is commonly found surrounding small orifices near the edges of lava lobes, as if it had been deposited from fluids escaping from beneath. This type of deposit seems to be primarily associated with the crestline of the mounds and may be the product of low-temperature diffuse venting.

Camera tows photographed tube worm colonies living on older lava between the new mounds along a hydrothermally active fracture system⁷, and also found tube worms living on new lava on mound 1 (Fig. 3) and mound 2. This gives an absolute constraint on the maximum time necessary for the biological colonization of new vent sites (2–6 years on mound 1). The hydrothermal vents on the new lavas may be due to leakage of warm fluids through the permeable new lava mounds from the buried fracture system.

Our interpretation is that all the fresh lava mounds were erupted during one rifting episode, either in one event or in several events over a period of years, analogous to Icelandic volcanic systems¹⁴. The eruption clearly occurred along a prominent fissure system that is now hydrothermally active⁷, and it is reasonable to assume that the lava mounds were fed from, and are underlain by, a recent dyke intrusion. Monitoring of sub-aerial volcanoes suggests that the emplacement of this dyke produced horizontal extension across the northern Cleft segment of the order of metres and therefore represents an episode of active sea-floor spreading. No seismicity was detected associated with this eruption (R. Dziak, personal communication); however, all seismic data for the northeast Pacific is from land-

based stations, and the detection threshold is m_b 4.0 (B. Presgrave, personal communication).

The Juan de Fuca megaplumes⁶ were observed in 1986 and 1987, within the time interval for the eruption. This coincidence in time supports earlier models that megaplumes may be caused by sea-floor spreading^{5–9} and the conclusion¹⁰ that the 1986 megaplume was followed by degassing of fresh magma into the hydrothermal system. Horizontal extension caused by dyke intrusion is a reasonable mechanism for a sudden increase in the permeability of the crust in the vicinity of a pre-existing hydrothermal system, leading to catastrophic release of hydrothermal fluids. Megaplumes may thus be indirect indicators of sea-floor spreading events. □

Received 19 November 1990; accepted 14 February 1991.

1. Macdonald, K. C., Haymon, R. & Shor, A. *Geology* **17**, 212–216 (1989).
2. Nishimura, C. E., Vogt, P. R., Smith, L. & Boyd, J. D. *EOS Trans. Am. geophys. Un.* **70**, 1301 (1989).
3. Bergman, E. A. & Solomon, S. C. *J. geophys. Res.* **95**, 4943–4965 (1990).
4. Shor, A. N., Nishimura, C. E., Czarnecki, M. & Vogt, P. R. *EOS Trans. Am. geophys. Un.* **71**, 1602 (1990).
5. Baker, E. T., Massoth, G. J. & Feely, R. A. *Nature* **329**, 149–151 (1987).
6. Baker, E. T. *et al.* *J. geophys. Res.* **94**, 9237–9250 (1989).
7. Embley, R. W., Chadwick, W. W. Jr, Perfit, M. R. & Baker, E. T. *Geology* (in the press).
8. Cann, J. R. & Strens, M. R. *J. geophys. Res.* **94**, 12277–12287 (1989).
9. Embley, R. W. & Chadwick, W. W. Jr *EOS Trans. Am. geophys. Un.* **71**, 259 (1990).
10. Baker, E. T. & Lupton, J. E. *Nature* **346**, 556–558 (1990).
11. Moore, R. B. *et al.* *J. Volcan. geotherm. Res.* **7**, 189–210 (1980).
12. Björnsson, A. *J. geophys. Res.* **90**, 10151–10162 (1985).
13. Sigurdsson, H. *Geol. Assoc. Canada Spec. Pap.* **34**, (eds Hall, H. C. & Fahrig, W. F.) 55–64 (Geological Association of Canada, St John's, Newfoundland, 1987).
14. Björnsson, A., Johnsen, G., Sigurdsson, S., Thorbergsson, G. & Tryggvason, E. *J. geophys. Res.* **84**, 3029–3038 (1979).

ACKNOWLEDGEMENTS. D. Clague and J. Morton provided helpful reviews. This research was supported by the NOAA VENTS Program and the NOAA Undersea Research Program. Authorship of this paper is alphabetical.

Late Quaternary extinct ungulates of East Africa and palaeoenvironmental implications

Curtis W. Marean* & Diane Gifford-Gonzalez†

* Department of Anthropology, State University of New York at Stony Brook, Stony Brook, New York 11794-4364, USA

† Board of Studies in Anthropology, University of California at Santa Cruz, Santa Cruz, California 95064, USA

UNGULATE communities of two East African savannas, the Serengeti and Athi-Kapiti Plains, are dominated by wildebeest (*Connochaetes taurinus*) supplemented by zebra (*Equus burchelli*), topi (*Damaliscus lunatus*), hartebeest (*Alcelaphus buselaphus*), buffalo (*Syncerus caffer*), eland (*Taurotragus oryx*) and gazelles (*Gazella granti* and *G. thomsoni*)^{1–3}. Before this research, little was known of East African large mammal communities in the Late Pleistocene and early to middle Holocene. We document an extinct impala-sized alcelaphine antelope that is numerically dominant in Late Pleistocene archaeofaunal assemblages from the Athi-Kapiti Plains. The extinct giant buffalo *Pelorovis antiquus* is present, and a number of arid-adapted regionally extinct species are common. The small alcelaphine is rare in northern Tanzania, but regionally extinct arid-adapted species are present in Late Pleistocene deposits. These data indicate that as recently as 12,000 years ago, the large mammal community structure of East African savannas was very different and dry grasslands and arid-adapted ungulates expanded at least as far south as northern Tanzania during the Last Glacial Maximum.

The largest samples of Late Pleistocene fossils come from archaeological sites at Lukenya Hill, a gneissic inselberg on the northern fringes of the Athi-Kapiti Plains in south-central Kenya. There are many Later Stone Age archaeological localities

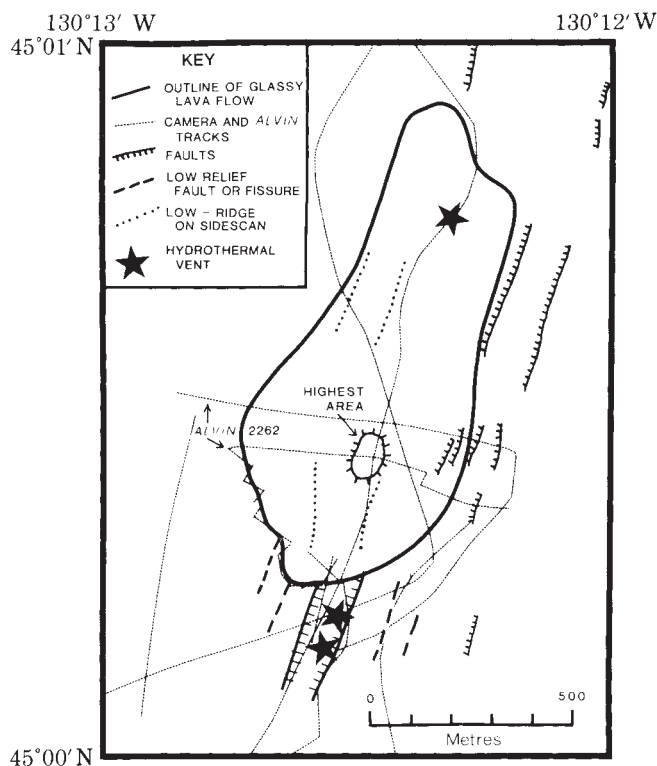


FIG. 3 Detailed map (outline shown in Fig. 2c) of mound 1 based on SeaMARC I sidescan sonar data, and the locations of flow contacts as surveyed by camera tows and Alvin dive 2262 (thin dashed lines). Note close correspondence with outline of mound 1 in Fig. 2c, derived independently from Sea Beam data.

TABLE 1 The occurrence of a socket for the lower second premolar in modern and fossil alcelaphines

Taxon	Present (both sides)	Present (one side)	Absent (all)	Sample size
Modern Specimens				
Hartebeest	100% (29)	0% (0)	0% (0)	29
Lichtenstein's hartebeest	100% (11)	0% (0)	0% (0)	11
Black wildebeest	0% (0)	0% (0)	100% (10)	10
Blue wildebeest	0% (0)	0% (0)	100% (22)	22
Bontebok/blesbok	60% (15)	4% (1)	36% (9)	25
Hunter's hartebeest	22% (2)	12% (1)	66% (6)	9
Topi	100% (30)	0% (0)	0% (0)	30
Fossil specimens				
Small alcelaphine	—	0% (0)	100% (5)	5
<i>D. agelaius</i>	—	0% (0)	100% (7)	7

Fossil samples are from Lukenya Hill, Gol Kopjes and Olduvai Gorge. Values are expressed as percentage of observed with sample size in parentheses. All fossil specimens were half mandibles so presence for 'both sides' is not relevant. All modern specimens for Tables 1 to 3 are wild specimens from collections in the National Museums of Kenya, the Field Museum in Chicago, and the American Museum in New York. Hartebeest, *A. buselaphus*; Lichtenstein's hartebeest, *A. lichtensteini*; black wildebeest, *C. gnou*; blue wildebeest, *C. taurinus*; bontebok/blesbok, *D. dorcas*; Hunter's hartebeest, *D. hunteri*; topi, *D. lunatus*. Fossil specimens were from: small alcelaphine, Lukenya Hill/Gol Kopjes; *D. agelaius*, Olduvai Gorge.

(about 40,000 yr BP to 1,000 yr BP) and four are reported here, numbered GvJm19, GvJm22, GvJm46, and GvJm62. GvJm19 and GvJm22 are rockshelters. At GvJm19 two stratigraphic components are dated by seven radiocarbon dates between 4,145 and 8,170 yr BP on the Holocene component⁴ and one date of 13,705 yr BP on the Late Pleistocene component. At GvJm22 two main stratigraphic components are dated by three late Holocene radiocarbon dates, and four Late Pleistocene radiocarbon dates between 13,730 and 17,700 yr BP (ref. 5). GvJm46 and GvJm62 are deeply stratified open-air sites that have yielded five Late Pleistocene radiocarbon dates between 12,195 and 21,535 yr BP (refs 4, 6). The radiocarbon dates indicate that most Lukenya Hill Late Pleistocene deposits date to the Last Glacial Maximum. The following numbers of bones and dentitions are

identifiable at least to family and body size: GvJm19, over 1,200; GvJm22, over 3,000; GvJm46, over 4,000; and GvJm62, about 500. The small alcelaphine is the dominant ungulate in all Lukenya Hill Late Pleistocene deposits but is absent from Holocene deposits. Other significant species in Late Pleistocene levels include oryx (*Oryx gazella*), Grevy's zebra (*Equus grevyi*), and the extinct giant buffalo *Pelorovis antiquus*.

The same small alcelaphine is rare in two sites in northern Tanzania: Kiseke II Rockshelter and Gol Kopjes. Kiseke II is a large, deeply stratified rockshelter about 70 km south of Lake Manyara. Its deposits span the Late Pleistocene and Holocene^{7,8}. Approximately 3,000 elements are identifiable to family and body size. In Late Pleistocene deposits alcelaphines and warthog (*Phacochoerus aethiopicus*) predominate; a notable occurrence is the historically absent Grevy's zebra. The small alcelaphine is represented by only nine isolated teeth. The Gol Kopje site, HcJel, is a large stratified open site in the Serengeti Plains. The deposit of 0.80 m dates from the present to at least the middle Holocene. Bone apatite dates of 7,310 and 6,920 yr BP occur at 40–50 cm; deeper levels were not dated⁹. The assemblage has over 6,000 identifiable elements. Burchell's zebra, wildebeest, Thomson's gazelle, Grant's gazelle and topi dominate, whereas the small alcelaphine is rare in all levels. HcJel possibly documents the small alcelaphine's latest occurrence but dating is problematic.

Virtually all of the diagnostic materials from the reported sites are teeth or fragmented postcrania. Many of the alcelaphine teeth are smaller and morphologically distinct from extant East African alcelaphines. The small alcelaphine molar occlusal morphology is less complex than that of extant East African alcelaphines (Fig. 1); the internal enamel cavities are extremely simple and the buccal walls (hypoconid and protoconid) of the lower molars are more rounded. There are no mandibular P₂s in the assemblages, and the five available mandibles lack a socket for P₂. Only wildebeest and bontebok/blesbok consistently lack P₂ (see Table 1). The fossil teeth from Lukenya Hill and Gol Kopjes are significantly smaller ($P < 0.001$, Mann-Whitney *U* test, middle-wear specimens only; Tables 2 and 3) than those of modern species except bontebok/blesbok. M³s were first classified into wear states as defined by Klein and Cruz-Urbe¹⁰ to minimize ontogenetic affects¹¹. The coefficients of variation suggest homogeneous samples¹² for all taxa.

The only extant East African alcelaphines that lack P₂ are blue wildebeest and occasionally Hunter's hartebeest, but the fossil specimens are too small and morphologically distinct to

TABLE 2 Length of occlusal surface

Taxon	Mean	Standard deviation	Coefficient of variation	Range	Sample size	M ₃ HI
Modern specimens						
Hartebeest	22.0	1.4	6.36	19.8–25.5	14	5.23
Lichtenstein's hartebeest	24.9	1.7	6.83	23.1–27.7	6	—
Black wildebeest	22.3	1.8	8.07	20.2–25.3	7	4.75
Blue wildebeest	27.3	1.8	6.59	23.9–30.3	22	4.94
Bontebok/blesbok	18.9	1.2	6.35	16.7–21.4	14	4.76
Hunter's hartebeest	22.6	1.2	5.31	20.6–24.3	6	4.14
Topi	20.8	1.6	7.69	18.5–24.0	14	5.10
Fossil specimens						
Small alcelaphine	18.6	1.0	5.37	17.0–20.8	40	5.74
<i>A. buselaphus/D. lunatus</i>	21.6	1.7	7.87	19.5–24.4	5	—
<i>C. taurinus</i>	25.5	1.1	4.31	24.0–27.1	4	—
<i>D. agelaius</i>	19.4	1.2	6.19	17.8–21.7	18	—

Summary statistics in millimeters for greatest length of the occlusal surface of middle-wear upper third molars, and the hypsodonty index (HI) of lower third molars, for modern alcelaphines and fossil alcelaphines from Lukenya Hill, Gol Kopjes and Olduvai Gorge. Sample sizes for modern taxa represent different individuals, whereas fossil samples comprise all measurable teeth. Hypsodonty index of modern specimens is from Janis²² and represents from one to four individuals for each species and three individuals in the case of the fossil small alcelaphine. HI standard deviation of fossil small alcelaphine was 0.02 and unknown for the data from Janis²². Modern specimens as in legend to Table 1; Fossil specimens are from: *A. buselaphus/D. lunatus* and *C. taurinus*, Lukenya Hill; small alcelaphine and *D. agelaius*, as in legend to Table 1.

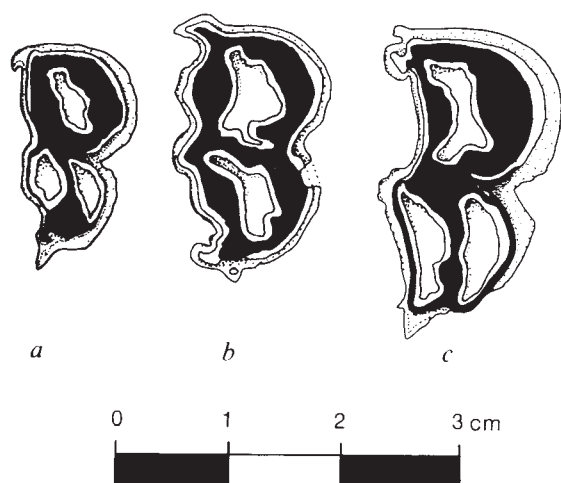


FIG. 1 Occlusal view of right middle-wear M^3 of small fossil alcelaphine antelope (a), fossil *Alcelaphus buselaphus* (b), and fossil *Connochaetes taurinus* (c); all specimens are Late Pleistocene from Lukenya Hill.

be either of these species. Bontebok/blesbok commonly lacks P_2 , shares similar morphological features with the small fossil alcelaphine, and overlaps in size. But the molars of the extinct small alcelaphine are more hypsodont than all extant alcelaphines (Table 2). Also, the P_3 of the Gol Kopjes specimens has a less inflated paraconid and more elongate metaconid-entoconid area than a sample ($n=6$) of modern South African bontebok/blesbok specimens examined by one of us (D.G.-G.).

In some features the teeth resemble the small fossil alcelaphine *Damaliscus agelaius*, well documented at Olduvai Gorge Beds III/IV¹³, but they are significantly smaller (Tables 2 and 3). *D. agelaius* lacks P_2 and it is possible that the small alcelaphine reported here is a smaller form of *D. agelaius*. An alcelaphine smaller than blue wildebeest is reported from Late Pleistocene deposits at Mumba-Höhle in northern Tanzania¹⁴, but it is larger than the form reported here. Dentitions of a small alcelaphine have been found at the Middle Pleistocene locality of Isenya, Kajiado District of Kenya (J. P. Brugal, personal communication). There is a damaged cranium and a nearly complete mandible from a small alcelaphine from the Middle Pleistocene locality of Lainyamok¹⁵, and bontebok/blesbok has been reported from Late Pleistocene deposits in Zimbabwe, outside its modern range¹⁶. All these finds of small alcelaphines of Middle and Upper Pleistocene age make the picture of alcelaphine evolution complex but fascinating. Until more complete fossil remains are found and dating problems resolved, we cannot with confidence attribute to a species the small alcelaphine antelope described

TABLE 3 Occlusal length comparison

Taxon	Small alcelaphine from Lukenya Hill/Gol Kopjes	<i>D. agelaius</i> from Olduvai Gorge
Modern specimens		
Hartebeest	5.3 ($P<0.001$)	4.0 ($P<0.001$)
Lichtenstein's hartebeest	3.9 ($P<0.001$)	3.6 ($P<0.001$)
Black wildebeest	4.2 ($P<0.001$)	3.4 ($P<0.001$)
Blue wildebeest	6.4 ($P<0.001$)	5.4 ($P<0.001$)
Bontebok/blesbok	1.0 ($P>0.05$)	1.0 ($P>0.05$)
Hunter's hartebeest	3.8 ($P<0.001$)	3.3 ($P<0.001$)
Topi	4.3 ($P<0.001$)	2.4 ($P<0.05$)
Fossil specimens		
Small alcelaphine	—	2.7 ($P<0.05$)
<i>D. agelaius</i>	2.7 ($P<0.05$)	—

Significance statistics for comparisons of middle-wear upper third molars occlusal length. Z values are from Mann-Whitney U Test. Probabilities are in parentheses adjacent to Z value and are two-tailed. Modern specimens and locations of fossil specimens as in legend to Table 1.

here, though we consider it improbable that it is bontebok/blesbok because of the morphological details discussed above.

In deposits dating to the Last Glacial Maximum at Lukenya Hill and Kisese II, the small alcelaphine is associated with two arid-adapted species, Grevy's zebra and oryx. Both species are absent from the modern Lukenya Hill region, and Grevy's zebra is absent from northern Tanzania. *Pelorovis antiquus*, present in the Lukenya Hill sites, is typically associated with faunal communities suggestive of arid habitats^{17,18}. These data indicate that the small alcelaphine also was arid-adapted. Grevy's zebra is currently found at and north of the equator. Lukenya Hill is at 1°30'S and Kisese II is at 4°50'S suggesting a substantial southern extension of dry grasslands during the Last Glacial Maximum. The small alcelaphine is the dominant bovid in the Lukenya Hill assemblages whereas hartebeest and wildebeest are extremely rare. These data suggest that the ungulate community structure of the Athi-Kapiti Plains, and other East African savannas, was much different during the Last Glacial Maximum and as recently as 12,000 years ago or later. More specifically, the numerical domination by large herds of wildebeest may be a recent phenomenon restricted to the Holocene. Many of the seasonally moist grasslands common to East Africa today which are necessary for the large wildebeest herds were probably more arid and characterized by dry grasses and scrub vegetation during the Last Glacial Maximum¹⁹. These results are consistent with hypothesized Last Glacial Maximum vegetation changes in East Africa²⁰ and also by recent estimates of mean annual rainfall decreases of 30% (ref. 21). □

Received 1 October 1990; accepted 12 February 1991.

1. Sinclair, A. R. E. in *Serengeti: Dynamics of an Ecosystem* (eds Sinclair, A. R. E. & Norton-Griffiths, M.) 1-30 (University of Chicago Press, Chicago, 1979).
2. Talbot, L. M. & Talbot, M. H. *The Wildebeest in Western Masailand* (The Wildlife Society, Monograph number 12, 1963).
3. Foster, J. B. & Coe, M. J. *J. Zool. Soc. Lond.* **155**, 413-425 (1968).
4. Nelson, C. M. & Kimengich, J. in *Origin and Early Development of Food Producing Cultures in North-Eastern Africa* (ed. Krzyzaniak, L.) 481-487 (Polish Academy of Sciences, Poznan, 1984).
5. Gramly, R. M. *Man* **11**, 319-344 (1976).
6. Miller, S. F. *Nyame Akuma* **14**, 31-34 (1979).
7. Inskip, R. in *Actes du IV Congrès Panafricain de Préhistoire* 249-256 (Tervuren, 1962).
8. Deacon, J. *Ann. Cape Prov. Mus.* **5**, 5-84 (1966).
9. Bower, J. R. F. & Smale, W. *The Longest Record: The Human Career in Africa* (Conf., Berkeley, 1986).
10. Klein, R. G. & Cruz-Uribe, K. *The Analysis of Animal Bones from Archaeological Sites* (University of Chicago, Chicago, 1984).
11. Vrba, E. *Ann. Transvaal Mus.* **15**, 287-352 (1973).
12. Simpson, G. C., Roe, A. & Lewontin, R. C. *Quantitative Zoology* (Harcourt, Brace & World, New York, 1960).
13. Gentry, A. W. & Gentry, A. *Bull. Br. Mus. nat. Hist. Geol.* **29**, 289-446 (1978).

14. Lehman, U. *Mitt. Geol. Staatsinst. Hamburg* **26**, 100-140 (1957).
15. Potts, R., Shipman, P. & Ingall, E. *J. hum. Evol.* **17**, 597-614 (1988).
16. Klein, R. G. *Occ. Papers Nat. Mus. Mon. Rhodesia* **4**, 74-80 (1978).
17. Klein, R. G. *Ann. S. Afr. Mus.* **81**, 233-283 (1980).
18. Vrba, E. *Palaeont. Afr.* **26**, 33-46 (1987).
19. Marean, C. W. thesis, Univ. California at Berkeley (1990).
20. Hamilton, A. C. *Environmental History of East Africa* (Academic, New York, 1982).
21. Bonnefille, R. et al. *Nature* **346**, 347-349 (1990).
22. Janis, C. M. in *Teeth Revisited: Proceedings of the Vllth International Symposium on Dental Morphology* (eds Russell, D. E., Santoro, J.-P. & Sigogneau-Russell, D.) 367-387 (*Mémoires du Muséum National d'Histoire Naturelle, Série C, tome 53*, Paris, 1988).

ACKNOWLEDGEMENTS. We thank A. W. Gentry, F. Grine, R. G. Klein, L. Martin, L. Orloff, G. Richards and T. D. White for helpful comments. E. Vrba for consulting with D.G.G., the Office of the President of the Republic of Kenya, the National Museums of Kenya, and the Tanzanian Department of Antiquities for permission to study these materials. C.W.M. was funded by Boise Fund, LSB Leakey Foundation, National Science Foundation, and Sigma Xi. D.G.G. was supported by National Science Foundation grants to J.R.F. Bower, and by funding from the National Science Foundation and faculty research funds from the University of California, Santa Cruz.

No cost of echolocation for bats in flight

J. R. Speakman & P. A. Racey

Department of Zoology, University of Aberdeen, Aberdeen AB9 2TN, UK

ECHOLOCATION has evolved in relatively few animal species¹. One constraint may be the high cost of producing pulses, the echoes of which can be detected over useful distances². The energy cost of echolocation in a small (6 g) insectivorous bat, when hanging at rest, was recently measured at 0.067 Joules per pulse³, implying a mean cost for echolocation in flight of $9.5 \times$ basal metabolic rate (range 7 to 12 $\times$). Because flight is very costly⁴, whether the costs of echolocation and flying are additive is an important question. We measured the energy costs of flight in two species of small echolocating Microchiroptera using a novel combination of respirometry and doubly-labelled water⁵. Flight energy expenditure (adjusted for body mass) was not significantly different between echolocating bats and non-echolocating bats and birds. The low cost of echolocation for flying vertebrates may have been a significant factor favouring its evolution in these groups.

Two techniques have been used to estimate the energy costs of flight from respiratory gas exchange in bats and birds—respirometry and doubly-labelled water (DLW). Here we combined the two techniques by measuring the total energy expended by bats over an average period of 3.3 h using DLW. Approximately 0.5 h of this time was spent in free flight in a large room, whereas most of the rest was spent in a respirometry chamber. Assuming that non-flying time spent out of the chamber involved expenditure of energy at the same rate as when in the chamber, we estimated the non-flight costs and hence measured the flight costs by the difference.

Of 28 bats used, 25 were pipistrelles *Pipistrellus pipistrellus* (5.4 to 9.4 g) and three were brown long-eared bats *Plecotus auritus* (6.7 to 8.4 g). Bats were labelled with heavy oxygen and deuterium and placed in a respirometer for 90 min, after which blood samples were taken. The bats were then allowed to fly freely in a large darkened room (0.01 lux) for 60 min, followed by between 90 and 240 min of respirometry. A second blood sample was then taken. Five individuals flew for less than 600 s each, which was less than 10% of the time not measured by respirometry. Of the remainder, the time spent flying varied between 40 and 96% of the free-flight period (1,160 to 3,110 s). We failed to obtain sufficient blood at the second sampling for isotope analysis in six individuals.

For the five bats which flew for less than 600 s we evaluated energy expenditure during the entire time that they were out of the respirometry chambers. Because the time spent in flight by these bats was low, their estimated energy expenditure should be approximately the same as that measured at rest in the chambers. Therefore, these five estimates acted as controls. Flight cost estimates were made for 14 *P. pipistrellus* and 3 *P. auritus*.

The mean pre-flight energy expenditure over the 90 min prior to flight did not vary with time (Fig. 1a) and averaged 0.55 W (s.d. = 0.0318, $n = 90$). Mean post-flight energy expenditure initially declined, from a level immediately after flight which was similar to the pre-flight level, to a minimum an hour later. Energy expenditure remained at this low level for a further 100 min. Only six individuals were monitored for longer than 160 min. These followed the same pattern as the other bats over the first 160 min and thereafter their energy expenditure increased to the pre-flight level (Fig. 1b). Over the first 160 min of the post-flight period the average energy expenditure was significantly lower ($t = 3.69$, d.f. = 248, $P < 0.001$) than the energy expenditure over the 90 min pre-flight in the same individuals (mean post-flight expenditure 0.358 W, s.d. = 0.052, $n = 160$). The difference between pre- and post-flight energy

expenditure in individual bats was not significantly correlated to either the time spent in flight, the energy expended in flight or body mass. Therefore the observed depression was probably not a compensatory mechanism to cover the energy cost of the flying.

The mean energy expenditure whilst out of the chamber for the five individuals which flew for less than 1,000 s was 0.56 W. This was evaluated by subtracting the energy expended within the chamber from the DLW estimate of total energy expenditure. During this time the bats spent >90% of their time at rest and their energy expenditure did not differ significantly from the mean pre-flight or immediate post-flight resting energy expenditure measured in the respirometry chamber.

For the 17 bats which flew for longer than 1,000 s there was no relationship between the flight energy expenditure and body mass (Fig. 2). The mean energy expenditure for the pipistrelles was 1.43 W (s.d. = 0.511, $n = 14$) and for the long-eared bats was 1.3 W (s.d. = 0.554, $n = 3$). These latter estimates do not differ significantly from the previous estimate for flight cost in *P. auritus*, evaluated by DLW alone, of 1.49 W (ref. 6).

The large individual variation in flight cost estimates (Fig. 2) was a consequence of precision error in the DLW technique. We have previously validated the DLW technique⁷ by comparison with respirometry in these species and shown that the mean deviation of the DLW estimates from simultaneous respirometry was 13.5%. Because we subtracted the calculated resting costs measured by respirometry from the DLW estimates this error was amplified. The bats flew for an average of 1,754 s,

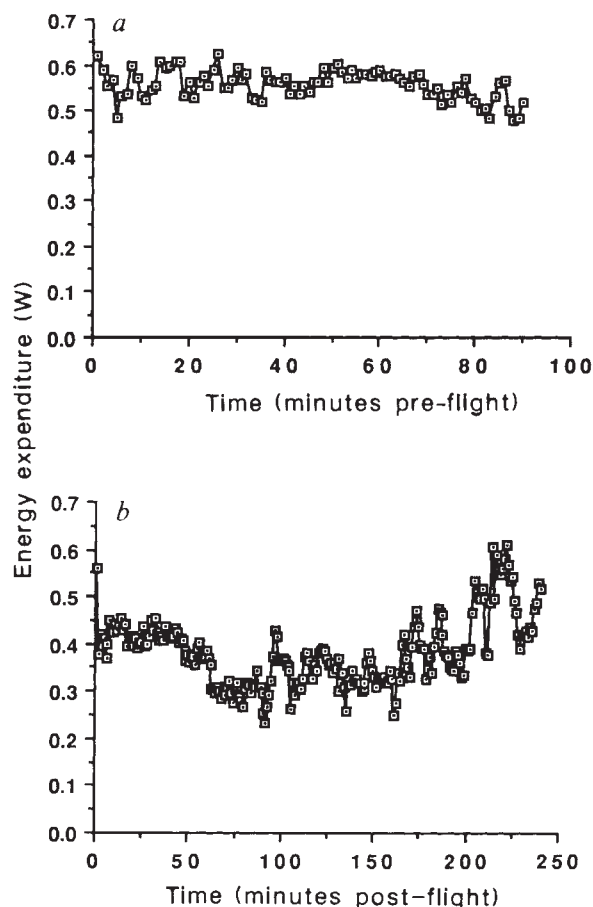


FIG. 1 a, Resting energy expenditure of bats averaged across 21 individuals measured over 90 min prior to flying for approximately half an hour and b, over 250 min following flight. During the post-flight phase the energy expenditure was depressed.

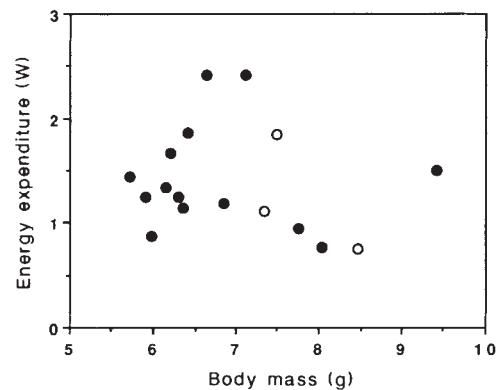


FIG. 2 Flight energy expenditure of small echolocating bats measured using a combination of DLW and respirometry, as a function of body mass. The closed symbols represent pipistrelle bats (*Pipistrellus pipistrellus*) and the open symbols brown long-eared bats (*Plecotus auritus*). Energy expended was measured using DLW over a period of 3.3 h, approximately 0.5 h of which was spent in flight and the majority of the remainder at rest in a respirometer. Carbon dioxide production was calculated using the Lifson and McClintock²⁷ equation with the pool size estimated from the dilution space of the ¹⁸oxygen. This was converted to energy expenditure using the mean RQ of the pre- and post-flight respirometry sessions. The error in converting carbon dioxide production to energy expenditure using a known RQ is less than 1%²⁸. We assume that the energy expenditure whilst out of the respirometry chamber, but not in flight, was at the same rate as the pre- and immediate post-flight resting rates. Therefore, the cost of flight was assessed from the difference between the DLW measure of total energy expenditure and the respirometry estimate of non-flying energy expenditure.

with a mean energy expenditure of 1.396 W. The rest time averaged 11,100 s, with a mean energy expenditure of 0.356 W and the expected amplification of variation was $\times 2.38$. The expected coefficient of variation in flight cost estimates due to analytical precision error was therefore 34.6%. The amplification of error masked any individual variation from biological sources. The use of DLW in combination with respirometry to estimate flight costs therefore provides no advantage over the use of DLW alone, when the time in flight is short relative to the time spent at rest, because it provides only a mean estimate across several individuals.

There is also an accuracy error in DLW, when compared with respirometry, which varies with the equation which is employed⁷. For the Lifson and McClintock equation, used to calculate CO₂ production here, we evaluated the mean accuracy error as +9.4% ($n = 9$)⁷. Because of amplification of error, this

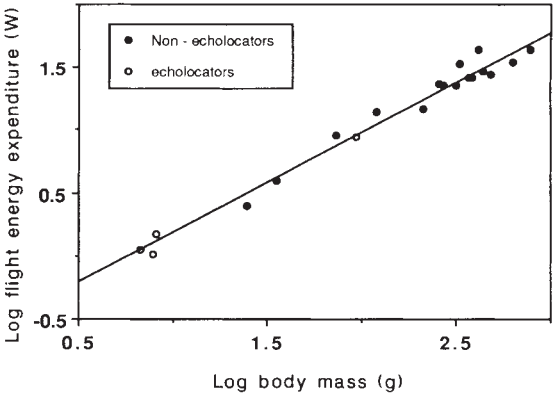
FIG. 3 Energy expended in flight of echolocating bats (open symbols) and non-echolocating birds and bats (closed symbols) plotted against body mass (g). The regression equations relating energy expenditure to body mass did not differ significantly between the two groups. For details of data used refer to Table 1. The best-fit equation for the combined data was $\log_{10}$ energy expenditure (W) = $-0.638 + 0.808 \log_{10}$ body mass (g); $r^2 = 0.981$.

TABLE 1 Estimates of energy expenditure for non-echolocating bats and birds, and for echolocating bats

Species	Mass (kg)	Energy expenditure (W)	n	Ref.	Technique*
Non-echolocating bats					
<i>Pteropus gouldii</i> (= <i>alecto</i>)	0.779	43.7	1	12	R
<i>Pteropus poliocephalus</i>	0.628	34.8	4	13, 14	R
<i>Hypsignathus monstrosus</i>	0.258	23.3	1	14	R
<i>Eidolon helvum</i>	0.315	22.4	1	14	R
Non-echolocating birds					
<i>Meliphaga virescens</i>	0.0243	2.46	?	15	R
<i>Melopsittacus undulatus</i>	0.035	3.93	1	16	R
<i>Sturnus vulgaris</i>	0.0728	9.15	6	17	R
<i>Falco spawerius</i>	0.120	13.8	3	18	R
<i>Falco tinnunculus</i>	0.213	14.6	6	19	R
<i>Corvus ossifragus</i>	0.275	22.8	2	20	R
<i>Larus atricilla</i>	0.370	26.2	2	21	R
<i>Columba livia</i>	0.384	25.9	4	22	D
<i>Columba livia</i>	0.442	29.5	6	23	R
<i>Columba livia</i>	0.412	43.7	6	24	D
<i>Columba livia</i>	0.330	34.0	6	4	R
<i>Corvus cryptoleucos</i>	0.480	27.8	2	25	R
Echolocating bats					
<i>Phyllostomas hastatus</i>	0.093	8.83	1	26	R
<i>Plecotus auritus</i>	0.008	1.35	3	6	D
<i>Pipistrellus pipistrellus</i>	0.0067	1.12	17	This study	R/D
<i>Plecotus auritus</i>	0.0077	1.02	3	This study	R/D

* R refers to measurements made by respirometry and D to measurements made by doubly-labelled water. Measurements here were made by a novel combination of the two techniques. We treated as independent data values generated for different groups of animals in independent studies.

indicates that the estimates of flight cost were 23.3% too large and in our previous study⁶, 9.4% too large. We revised our current and previous⁶ estimates in the light of the validation results⁷. All flight cost estimates for echolocating bats and non-echolocating bats and birds, using either respirometry alone, DLW alone, or DLW and respirometry combined are summarized in Table 1. We included only those birds for which estimates were made during sustained flights at the minimum power or maximum range speeds (for example, excluding hovering hummingbirds). We also excluded measures made on hirundines, swifts and terns, which have morphological adaptations to reduce flight costs⁸ which are not found in the echolocating bats under study. Birds which spend a significant portion of the time gliding were also excluded. The gradients and intercepts of the relationships between $\log_{10}$ flight energy expenditure and $\log_{10}$ body mass, for echolocating bats and non-echolocating birds



and bats combined were not significantly different (Fig. 3; echolocators gradient = 0.7932, non-echolocators gradient = 0.8, analysis of covariance (ANCOVA) gradients $F = 0.034$, $p = 0.954$, d.f. 1,20). Thus, it is clear that the estimates of flight cost in echolocating bats are not significantly greater than those for non-echolocating bats.

The absence of an energy cost for echolocation for flying animals may explain why echolocation systems are widespread in the Microchiroptera, but have evolved in very few terrestrial animals. Furthermore the high cost for terrestrial animals³ may explain why those systems that have evolved in terrestrial mammals involve only very weak short-range pulses⁹. Two alternative hypotheses may also explain the paucity of, and low intensity of, terrestrial echolocation systems. First, the echolocation pulses may reveal the whereabouts of the emitter to potential prey and predators. Alternatively, in a complex and cluttered terrestrial environment the emitter may be confused by strong reflections from very close large objects. Our data strongly support the energy cost hypothesis but cannot rule out these alternative hypotheses in the evolution of echolocation.

The close link between flight and reduced costs for echolocation raises the issues of why echolocation has evolved so infrequently amongst the birds and Megachiroptera, and why, when it has evolved in these animals, it is primarily used for gross navigation rather than prey detection. We suggest that the paucity of echolocation systems amongst these groups reflects a phylogenetic constraint on the development of the processing capacity for complex echolocation signals in animals which are already evolutionarily committed to a visual system. Vision is clearly the dominant system amongst birds. Nocturnal birds have larger cortex areas devoted to processing olfactory stimuli than diurnal birds¹⁰. As total brain size remains unaffected this suggests there is a trade off in the processing capacity allocated to the various senses. This may have prevented any species from

making a complete evolutionary change from one system (visual) to another (echolocation) because the intermediate steps would be selectively inferior to either of the pure systems. The recent suggestion that the megachiroptera have a primate ancestry¹¹ is consistent with this interpretation because primates also have well-developed vision. □

Received 15 October 1990; accepted 12 February 1991

1. Nactigall, P. E. & Moore, P. W. B. (eds) *Animal Sonar: Processes and Performance* (Plenum, New York, 1988).
2. Dawkins, R. *The Blind Watchmaker* (Longman, Bath, 1986).
3. Speakman, J. R., Anderson, M. E. & Racey, P. A. *J. comp. Physiol.* **165**, 679–685 (1989).
4. Rothe, H. J., Biesel, W. & Nactigall, W. *J. comp. Physiol.* **157**, 99–109 (1987).
5. Nagy, K. A. *Am. J. Physiol.* **238**, 466–473 (1980).
6. Racey, P. A. & Speakman, J. R. *Symp. zool. Soc. Lond.* **57**, 107–125 (1987).
7. Speakman, J. R. & Racey, P. A. *Physiol. Zool.* **61**, 514–526 (1988).
8. Flint, E. N. & Nagy, K. A. *Auk* **101**, 288–294 (1984).
9. Gould, E., Negus, N. C. & Novick, A. *J. exp. Zool.* **156**, 19–38 (1964).
10. Healy, S. & Guilford, T. *Evolution* **44**, 339–347 (1990).
11. Pettigrew, J. D. *Science* **231**, 1304–1306 (1986).
12. Thomas, S. P. *J. exp. Biol.* **94**, 231–250 (1981).
13. Carpenter, R. E. *J. exp. Biol.* **114**, 619–647 (1985).
14. Carpenter, R. E. *J. exp. Biol.* **120**, 79–103 (1986).
15. Collins, B. G. & Morellini, P. C. *Physiol. Zool.* **52**, 165–175 (1978).
16. Tucker, V. A. *J. exp. Biol.* **48**, 67–87 (1968).
17. Torre-Bueno, J. R. & LaRoche, J. *J. exp. Biol.* **75**, 223–229 (1978).
18. Gessamen, J. R. *Comp. Biochem. Physiol.* **65A**, 273–289 (1980).
19. Masman, D. thesis, Univ. Groningen (1984).
20. Bernstein, M. H., Thomas, S. P. & Schmidt-Nielsen, K. *J. exp. Biol.* **58**, 401–410 (1973).
21. Tucker, V. A. *Am. J. Physiol.* **222**, 237–245 (1972).
22. Lefebvre, E. A. *Auk* **81**, 403–416 (1964).
23. Butler, P. J., West, N. H. & Jones, D. R. *J. exp. Biol.* **77**, 7–26 (1977).
24. Gessamen, J. A. & Nagy, K. A. *Condor* **90**, 662–668 (1988).
25. Hudson, D. M. & Bernstein, M. H. *J. exp. Biol.* **103**, 121–130 (1973).
26. Thomas, S. P. *J. exp. Biol.* **63**, 273–293 (1975).
27. Lifson, N. & McClintock, R. *J. theor. Biol.* **12**, 46–74 (1966).
28. Gessamen, J. A. & Nagy, K. A. *Physiol. Zool.* **61**, 507–513 (1988).

ACKNOWLEDGEMENTS. We thank P. I. Webb and A. M. Burnett for assistance, A. E. Fallick and T. Donnelly of the Scottish Universities Research and Reactor Centre, East Kilbride, for help and advice, and J. P. Hayes, G. C. Hays, M. B. Fenton and J. M. V. Rayner for constructive comments on the manuscript. This work was supported by the NERC.

A B cell-deficient mouse by targeted disruption of the membrane exon of the immunoglobulin μ chain gene

Daisuke Kitamura, Jürgen Roes, Ralf Kühn & Klaus Rajewsky*

Institute for Genetics, University of Cologne, Weyertal 121, D-5000 Cologne 41, Germany

OF the various classes of antibodies that B lymphocytes can produce, class M (IgM) is the first to be expressed on the membrane of the developing cells. Pre-B cells, the precursors of B-lymphocytes, produce the heavy chain of IgM (μ chain), but not light chains¹. Recent data suggest that pre-B cells express μ chains on the membrane together with the 'surrogate' light chains $\lambda 5$ and VpreB (refs 2–7). This complex could control pre-B-cell differentiation, in particular the rearrangement of the light-chain genes⁸. We have now assessed the importance of the membrane form of the μ chain in B-cell development by generating mice lacking this chain. We disrupted one of the membrane exons of the gene encoding the μ -chain constant region by gene targeting⁹ in mouse embryonic stem cells¹⁰. From these cells we derived mice heterozygous or homozygous for the mutation. B-cell development in the heterozygous mice seemed to be normal, but in homozygous animals B cells were absent, their development already being arrested at the stage of pre-B-cell maturation.

The vector used for the disruption of one of the membrane exons (μM) (Fig. 1A) contains 9 kilobases (kb) of genomic DNA spanning exons 1 and 2 of μM and the first three exons of the constant (C) region of the δ gene. Close to the 5' boundary of the first exon of μM we introduced a translational stop codon and a *SalI* site into which a neomycin-resistance gene (*neo^r*) cassette¹¹ was inserted. At the 3' end of the genomic sequence we placed the herpes simplex virus thymidine kinase gene to permit selection against random integration¹².

Cells of the embryonic stem cell clone D3¹³ were transfected with the linearized vector by electroporation and selected by G418 and gancyclovir on feeder layers of STO fibroblasts¹⁴. Surviving colonies were screened for homologous recombinants using the polymerase chain reaction (PCR) (see legend to Fig. 1). PCR-positive clones were expanded and their identification as homologous recombinants verified by Southern blotting (Fig. 1B). From 3.4×10^7 transfected embryonic stem cells 1,870 colonies were resistant to G418 (determined by control plates), 230 were resistant to both G418 and gancyclovir and in six clones one of the two *C μ* genes in the genome was modified by homologous recombination with the vector without random integration. Thus, the frequency of gene targeting was 1/38 G418^r+GANC^r (G418 and gancyclovir-resistant respectively) colonies, which corresponds to 1/312 G418^r colonies or 1/5.7 $\times 10^6$ transfected cells.

The mutated clones were injected into blastocysts from C57BL/6 mice to generate chimaeric animals. As the D3 line is derived from an agouti mouse (strain 129/Sv), chimaeric mice could be identified by coat colour. Ten male chimaeras derived from four different mutated clones were mated to C57BL/6 females. One of these chimaeras, derived from clone 210, transmitted embryonic stem-derived chromosomes into the germ line as judged by the production of agouti offspring at a

* To whom correspondence should be addressed.

frequency of 1:6. Three of ten such agouti animals contained the mutated allele as shown by Southern blotting (Fig. 2b). These animals permitted a first analysis of the effect of the μ M mutation, designated μ MT, on B-cell development. Flow cytometric analysis of peripheral blood B cells (identified by the CD45R(B220) surface antigen¹⁵) demonstrated normal levels of such cells in the heterozygous animals as compared with (C57BL/6 $\times$ 129/Sv)F₁ controls, but all cells expressed IgM of C57BL/6 origin, that is of the *b* allotype. This is in contrast to the situation in the F₁ controls where half of the cells express IgM^b and half IgM^a (of 129/Sv origin; Fig. 3). Thus, the μ MT mutation is indeed correctly targeted to the *C μ* locus, preventing the expression of the membrane form of the μ chain from the targeted allele.

Studies with Abelson virus-transformed pre-B cell lines⁸ and immunoglobulin transgenic mice¹⁶ have suggested that the μ chain of membrane-bound IgM (the μ _m chain) plays a crucial role in mediating allelic exclusion, that is, the expression in a single B cell, of only one allele at the heavy-chain loci. The

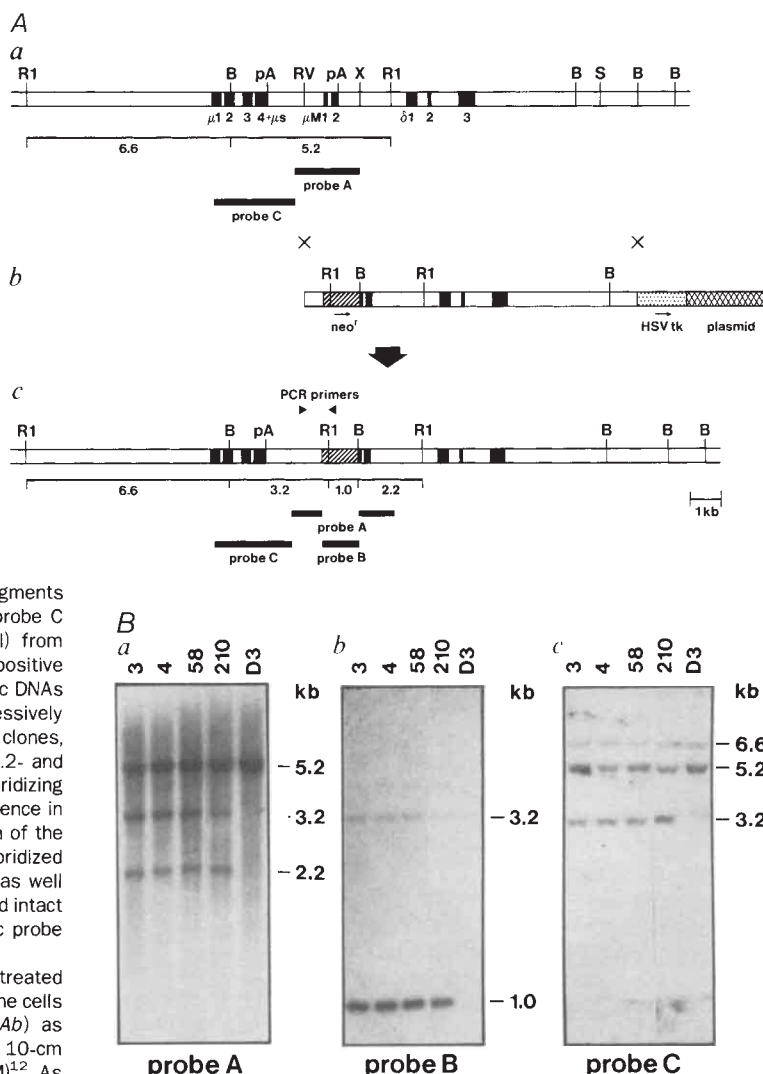
heterozygous mice depicted in Fig. 3 allow this hypothesis to be tested directly: if true, then productive rearrangement of the targeted locus should not preclude productive rearrangement on the other homologue. We indeed found that a small percentage of the B cells in these animals carry a-allotype antibodies of class D on the surface and that low levels of IgM^a can be detected in the serum (data not shown).

The effect of the μ MT mutation on the development of B lineage cells can best be studied in homozygous mutant mice. Such animals (14.2 and 14.3; Fig. 2c) were generated by intercrossing animals heterozygous for the mutation (G5 end G7; Fig. 2b). At the age of 4 weeks, cells from blood, spleen and bone marrow of μ MT/ μ MT animals were analysed for the presence of B lineage cells by flow cytometry. Cells from heterozygous (μ MT/+) and normal (+/+) littermates served as controls (Fig. 4). As expected, peripheral blood lymphocytes (Fig. 4a) and spleen cells (Fig. 4b) of the homozygous mutant mice lacked CD45R(B220)^{bright} cells, indicating the absence of mature B cells. We detected neither cells expressing IgM or IgD

FIG. 1 A, Strategy for the disruption of the membrane exon of the *C μ* gene. a, Genomic structure of the *a* allele of the *C μ* -*C δ* gene locus. Exons are represented by black boxes. pA, polyadenylation sites for the secretory (μ _s) and the membrane (μ _m) form of the μ chain²². The lengths of diagnostic restriction fragments and location of probes used for Southern blot analysis are shown. RI, *EcoRI*; B, *Bam*HI; RV, *EcoRV*; X, *Xho*I; S, *Sal*I; b, Targeting vector containing a 9-kb *EcoRV*-*Sal*I fragment of the *C μ* -*C δ* locus. The third codon of membrane exon 1 (μ M1 in a) was changed to TAG followed by an insertion of TCGAC to create a *Sal*I site by site directed mutagenesis. Five base pairs (bp) downstream of the *Sal*I site, 7 bp were accidentally deleted. An *Xho*I-*Sal*I fragment (*neo*^r) of pMC1Neo-polyA¹¹(PMC1POLA, STRATAGENE) was inserted into the new *Sal*I site. The *EcoRV*-*Sal*I fragment carrying the *neo*^r gene was inserted into an *Xho*I site of pLC19R/MC1-TK¹² (gift of K. Thomas, S. Mansour and M. Capecchi) which contains the HSV *tk* gene. The vector was linearized by *Cla*I before transfection. c, Predicted structure of the targeted locus. Triangles indicate the primers used for PCR assays. The sequence of the 5' primer (5'-CTCTGTAACCACTCACCACC-3') is located 96 bp upstream of the *EcoRV* site shown in a. The sequence of the 3' primer (5'-CCTGCGTGCAATCCATCTTG-3') is located 320 bp downstream of the *Xho*I site in pMC1Neo-polyA. The lengths of diagnostic restriction fragments and hybridization probes are indicated. Probe A (*Nco*I-*Xho*I) and probe C (*Nco*I-*Nco*I) are derived from the *C μ* locus, probe B (*Xho*I-*Sal*I) from pMC1Neo-polyA. B, Southern blot analysis of D3 cells and PCR-positive transfectants (four clones—3, 4, 58 and 210—are shown). Genomic DNAs were digested by *Bam*HI and *Eco*RI. The filter was hybridized successively by the three probes indicated below each panel. a, In the PCR-positive clones, probe A hybridized to a wild-type 5.2-kb fragment as well as to 3.2- and 2.2-kb fragments in the targeted locus, as expected. b, Probe B hybridizing to 3.2- and 1.0-kb fragments shows the presence of the *neo*^r sequence in the locus. The faint 4.2-kb band is probably due to partial digestion of the DNA. c, As predicted, probe C, located outside the targeting vector, hybridized in the PCR-positive clones to the same 3.2-kb fragment as above as well as to the wild type fragments (6.6 and 5.2 kb). The *C δ* locus remained intact in the mutated allele as shown by hybridization with a *C δ* -specific probe (data not shown).

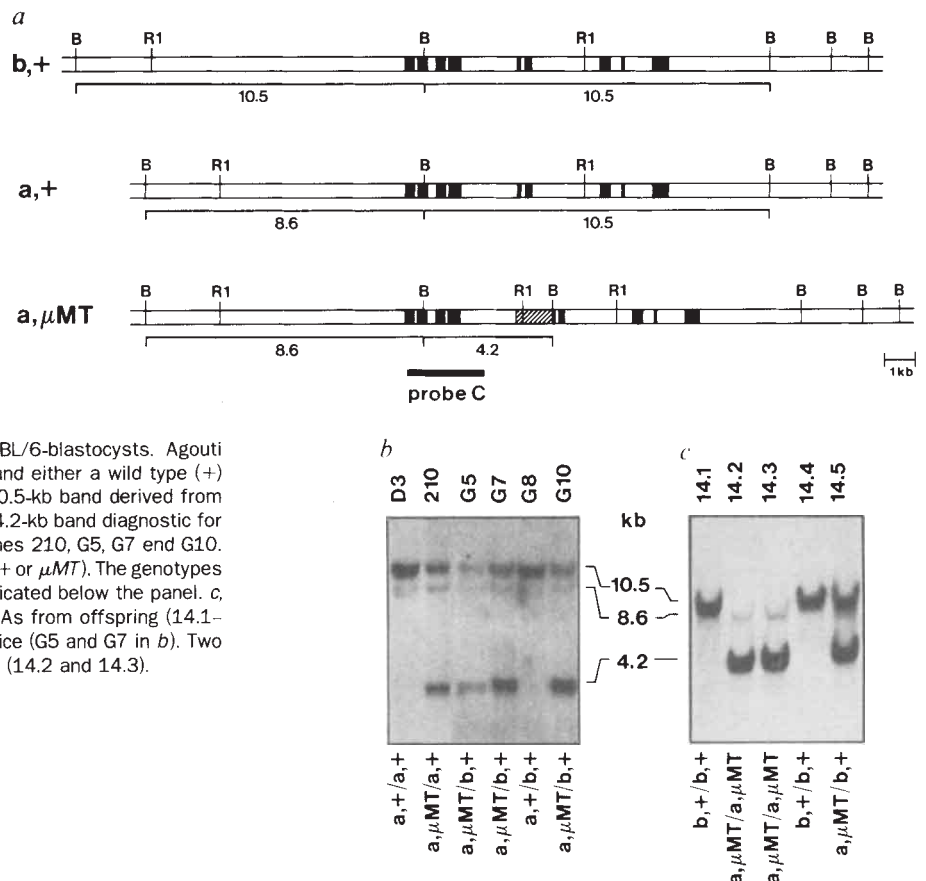
METHODS. D3 cells¹³ (gift of E. Wagner) were grown on mitomycin C-treated G418-resistant STO fibroblasts¹⁴ (gift of C. Kappen and F. Ruddle). The cells were electroporated with 20 μ g ml⁻¹ of the linearized vector (Ab) as described¹¹. The cells were plated at a density of 6 $\times$ 10⁶ cells per 10-cm feeder plate and selected with G418 (200 μ g ml⁻¹) and GANC (2 μ M)¹². As a control, 2 $\times$ 10⁶ transfectants were selected in G418 only, and gave rise to 110 G418^r colonies. After 11–14 days, G418^r-GANC^r colonies were screened for homologous recombination by PCR analysis as follows: the colonies were picked and trypsinized in multi-well plates individually. Half of the cells from each colony were cultured on multi-well plates. The other cells (8–12 colonies) were pooled for PCR analysis. Pooled samples were treated with proteinase K (ref. 23) and gene amplification was done in 50 μ l

10 mM Tris-HCl, pH 8.3, 50 mM KCl, 0.01% gelatine, 3 mM MgCl₂, 0.1 μ M primer (Ac), 0.4 mM dNTPs and 2.5 U *Taq* polymerase for 40 cycles (1.5 min at 95 °C, 1 min at 65 °C and 2 min at 72 °C). Samples containing targeted cells gave a 1-kb amplified fragment. The clones from PCR-positive pools were then analysed individually and positive clones were expanded and analysed by Southern blotting (B).



10 mM Tris-HCl, pH 8.3, 50 mM KCl, 0.01% gelatine, 3 mM MgCl₂, 0.1 μ M primer (Ac), 0.4 mM dNTPs and 2.5 U *Taq* polymerase for 40 cycles (1.5 min at 95 °C, 1 min at 65 °C and 2 min at 72 °C). Samples containing targeted cells gave a 1-kb amplified fragment. The clones from PCR-positive pools were then analysed individually and positive clones were expanded and analysed by Southern blotting (B).

FIG. 2 *a*, Genomic structure of the $C\mu$ - $C\delta$ loci of the *a* and *b* alleles and of the μMT mutant with predicted sizes of the fragments detected by probe C in Southern blot analysis (*b* and *c*). The Igh^b wild type allele (*b*, +; top) is from C57BL/6. The Igh^a wild type (*a*, +; middle) and mutated ($a, \mu MT$; bottom) alleles are from the mutated ES clones. *b*, Southern blot analysis showing germ-line transmission of the μMT mutation. DNAs from D3 cells (D3), a mutated D3-clone (210, Fig. 1*B*) and tails of agouti offspring (G5, G7, G8 and G10) were digested by *Bam*HI and hybridized with probe C. The offspring was from a C57BL/6 female mated to a chimaeric male which had been generated by the injection of clone 210 into C57BL/6-blastocysts. Agouti offspring carry an Igh^b allele of C57BL/6 origin and either a wild type (+) or a mutated (μMT) Igh^a allele of D3 origin. A 10.5-kb band derived from Igh^b or wild type Igh^a was present in all lanes. A 4.2-kb band diagnostic for the mutated allele ($Igh^a, \mu MT$) was present in lanes 210, G5, G7 and G10. A weak 8.6-kb band is derived from the Igh^a allele (+ or μMT). The genotypes of the $C\mu$ loci determined by the analysis are indicated below the panel. *c*, Southern blot analysis of *Bam*HI-digested tail-DNAs from offspring (14.1–14.5) of an intercross of heterozygous mutant mice (G5 and G7 in *b*). Two animals were homozygous for the μMT mutation (14.2 and 14.3).



on the surface nor B cells in the peritoneal cavity (data not shown). In addition, the homozygous mutant mice had no detectable IgM in the serum ($<0.1 \mu\text{g ml}^{-1}$), compared with $\sim 600 \mu\text{g}$ IgM per ml in the controls (data not shown). In contrast, $\mu MT/\mu MT$ mice generate substantial numbers of T cells, as demonstrated by staining with anti-Thy-1 (Fig. 4*a*) or anti-CD3 (Fig. 4*b*) antibodies.

Which stage of B cell development is affected by the μMT mutation? B cells are continuously generated from stem cells in the bone marrow. The latter cells first differentiate into pre-B cells, detectable by flow cytometry as $CD45R(B220)^{\text{dull}}$, surface

$IgM^{-}(sIgM^{-})$ cells¹⁷. Through rearrangement of the L chain genes, the pre-B cells give rise to $CD45R(B220)^{\text{dull}}$, $sIgM^{+}$ B cells^{1,17}. In the bone marrow of the homozygous mutant mice $sIgM^{+}$ cells were absent (Fig. 4*c*). By contrast, the frequency of cells of the pre-B cell phenotype was roughly the same as in the controls, although the level of $CD45R(B220)$ expression in these cells seemed slightly lower. Pre-B cells go through an early phase of proliferation as large cells from which small, non-dividing pre-B cells originate^{18–20}. Analysis of the $CD45R(B220)^{\text{dull}}$ cells in the homozygous mutant mice by forward light scatter showed that they were mainly large, in contrast

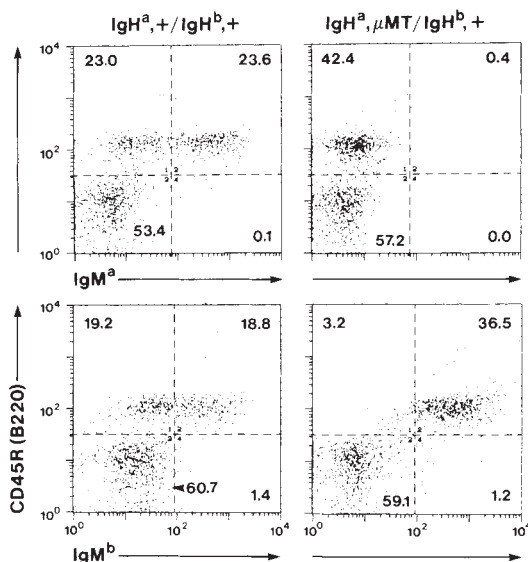
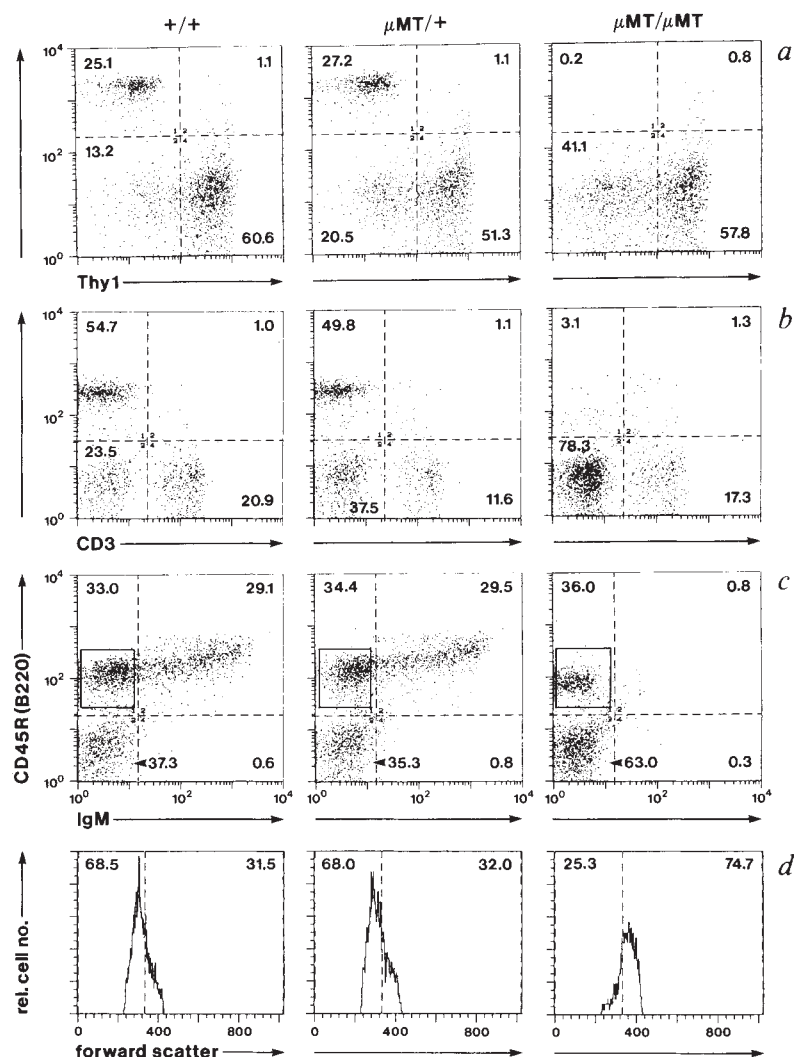


FIG. 3 Flow cytometric analysis (FACSscan, Becton-Dickinson) of peripheral blood lymphocytes (PBL) from 3-month-old F_1 mice carrying ($Igh^a, \mu MT/Igh^b, +$; mouse G7 in Fig. 2*b*) or not carrying ($Igh^a, +/Igh^b, +$; mouse G8 in Fig. 2*b*) the μMT mutation. Peripheral blood lymphocytes were purified on a Ficoll-gradient and stained with phycoerythrin (PE)-conjugated monoclonal antibody RA3-6B2²⁴ (anti- $CD45R(B220)$) and fluorescein isothiocyanate (FITC)-conjugated monoclonal antibodies, RS3.1²⁵ (anti- IgM^a) or MB86²⁶ (anti- IgM^b). Dots and percentages in the fluorescence windows refer to cells in the lymphocyte gate as defined by light scatter¹⁷. The vertical and horizontal axes show intensity of red (PE) and green (FITC) fluorescence, respectively.

FIG. 4 Flow cytometric analysis of lymphocytes from homozygous mutant ($\mu MT/\mu MT$; 14.3 in Fig. 2c), normal (+/+; 14.4 in Fig. 2c) and heterozygous mutant ($\mu MT/+$; 14.5 in Fig. 2c) littermates. *a*, PBLs, *b*, spleen cells or *c*, bone marrow cells from 4-week-old mice were stained with monoclonal antibody RA3-3A1¹⁵-PE (anti-CD45R(B220)), and mAbs CFO 1²⁷-FITC (anti-Thy 1, *a*) or 145-2C11²⁸-FITC (anti-CD3, *b*) or R33-24-12²⁹-FITC (anti-IgM, *c*). Dead cells were excluded by propidium-iodide staining in *b* and *c*. (See also legend of Fig. 3.) The boxes in *c* define those cell populations reanalysed for cell size by forward light scatter as displayed in *d* in the form of histograms.



to the pre-B cells in controls, of which only a small proportion were large (Fig. 4d). Therefore, the μMT mutation seems to arrest B-cell differentiation at the pre-B cell stage, presumably close to, or at the point of, transition from large to small pre-B cell. But, it cannot be excluded that the CD45R(B220)^{du} cells in the bone marrow of the homozygous mutant mice do not represent B-lineage cells; in this case the μMT mutation would prevent pre-B cell generation altogether.

The present work established that B-cell development is

dependent on the expression of the membrane form of the μ chain by the pre-B cell stage. This implies a function for this protein before L chain expression and supports the idea that μ_m together with 'surrogate' L chains forms a membrane receptor through which pre-B-cell differentiation is controlled. Furthermore, the μMT mouse mutant both allows testing of the hypothesis that allelic exclusion is mediated through μ_m expression and offers itself as a model of an immunodeficiency selectively affecting B-cell development, as occurs in humans²¹. □

Received 14 January; accepted 1 February 1991.

- Cooper, M. D. & Burrows, P. D. in *Immunoglobulin Genes* (eds Honjo, T., Alt, F. W. & Rabbitts, T. H.) 1-21 (Academic, London, 1989).
- Pillai, P. & Baltimore, D. *Nature* **329**, 172-174 (1987).
- Kerr, W. G., Cooper, M. D., Feng, L. & Hendershot, L. M. *Int. Immun.* **1**, 355-361 (1989).
- Takemori, T. et al. *EMBO J.* **9**, 2493-2500 (1990).
- Tsubata, T. & Reth, M. *J. exp. Med.* **172**, 973-976 (1990).
- Sakaguchi, N. & Melchers, F. *Nature* **324**, 579-582 (1986).
- Kudo, A. & Melchers, F. *EMBO J.* **6**, 2267-2272 (1987).
- Reth, M., Petrac, E., Wiese, P., Lobel, L. & Alt, F. W. *EMBO J.* **6**, 3299-3305 (1987).
- Capecci, M. R. *Science* **244**, 1288-1292 (1989).
- Evans, M. J. & Kaufman, M. H. *Nature* **292**, 154-156 (1981).
- Thomas, K. R. & Capecci, M. R. *Cell* **51**, 503-512 (1987).
- Mansour, S. L., Thomas, K. R. & Capecci, M. R. *Nature* **336**, 348-352 (1988).
- Doetschmann, T. C., Eistetter, M., Katz, M., Schmidt, W. & Kemler, R. *J. Embryol. exp. Morph.* **87**, 27-45 (1985).
- Robertson, E. J. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach* (ed. Robertson, E. J.) 71-112 (IRL, Oxford, 1987).
- Coffman, R. L. & Weissman, I. L. *Nature* **289**, 681-683 (1981).
- Manz, J., Denis, K., Witte, O., Brinster, R. & Storb, U. *J. exp. Med.* **168**, 1363-1381 (1988).

- Förster, I., Vieira, P. & Rajewsky, K. *Int. Immun.* **1**, 321-331 (1989).
- Opstelten, D. & Osmond, D. G. *J. Immun.* **131**, 2635-2640 (1983).
- Coffman, R. L. & Weissman, I. L. *J. molec. cell. Immun.* **1**, 31-38 (1983).
- Osmond, D. G. *Immunol. Rev.* **93**, 103-124 (1986).
- Cooper, M. D. & Butler, J. L. in *Fundamental Immunology* 2nd edn (ed. Paul, W. E.) 1033-1058 (Raven, New York, 1989).
- Guise, J. W., Galli, G., Nevins, J. R. & Tucker, P. W. in *Immunoglobulin Genes* (eds Honjo, T., Alt, F. W. & Rabbitts, T. H.) 275-301 (Academic, London, 1989).
- Kim, H.-S. & Smithies, O. *Nucleic Acids Res.* **16**, 8887-8903 (1988).
- Coffman, R. L. *Immunol. Rev.* **69**, 5-23 (1982).
- Schüppel, R., Wilke, J. & Weiler, E. *Eur. J. Immunol.* **17**, 739-741 (1987).
- Nishikawa, S., Sasaki, Y., Kina, T., Amagi, T. & Katsura, Y. *Immunogenetics* **23**, 137-139 (1986).
- Opitz, H. G. et al. *Immunobiology* **160**, 438-453 (1982).
- Leo, O., Foo, M., Sachs, D. H., Samelson, L. E. & Bluestone, J. A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1374-1378 (1987).
- Grützmann, R. thesis, Univ. Cologne (1981).

ACKNOWLEDGEMENTS. We thank W. Müller for reagents and advice, U. Ringeisen for artwork and E. Siegmund and G. Schmall for typing the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft through SFB 243 and the Fazit Foundation. D. Kitamura was supported by a fellowship from the Alexander von Humboldt Foundation, J. Roes from the Fonds der Chemie.

Mutational hotspot in the p53 gene in human hepatocellular carcinomas

I. C. Hsu*, R. A. Metcalf†, T. Sun‡, J. A. Welsh†, N. J. Wang§ & C. C. Harris†||

* Department of Pathology, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA

† Laboratory of Human Carcinogenesis, Division of Cancer Etiology, National Cancer Institute, Bethesda, Maryland 20892, USA

‡ Cancer Institute, C.A.M.S., Beijing, People's Republic of China

§ Qidong Liver Cancer Institute, Jiangsu, People's Republic of China

HUMAN hepatocellular carcinomas (HCC) from patients in Qidong, an area of high incidence in China, in which both hepatitis B virus and aflatoxin B₁ are risk factors¹, were analysed for mutations in p53, a putative tumour-suppressor gene. Eight of the 16 HCC had a point mutation at the third base position of codon 249. The G→T transversion in seven HCC DNA samples and the G→C transversion in the other HCC are consistent with mutations caused by aflatoxin B₁ in mutagenesis experiments^{2,3}. No mutations were found in exons 5, 6, 8 or the remainder of exon 7. These results contrast with p53 mutations previously reported in carcinomas and sarcomas of human lung, colon, oesophagus and breast; these are primarily scattered over four of the five evolutionarily conserved domains, which include codon 249 (refs 4–9). We suggest that the mutant p53 protein may be responsible for a selective clonal expansion of hepatocytes during carcinogenesis.

The p53 gene is frequently mutated in diverse human cancers^{4–9}. Germ-line p53 gene mutations are closely associated with the familial Li–Fraumeni syndrome in which multiple forms of cancer develop^{10,11}. Certain domains of p53 are highly conserved, reflecting its functional importance and selection in the p53 protein¹²; these domains and exons 5–8 are frequently the sites of mutation^{4–9}. We have therefore sequenced exons 5, 6, 7 and 8 and their flanking splicing regions after amplification by polymerase chain reaction (PCR), using liver tumour samples from 16 patients from Qidong with primary hepatocellular carcinomas. Sections of the p53 gene containing exons 5, 6, 7 and 8 and nearby intron regions were amplified from tumour genomic DNA. (When the tumour contained a mutation, non-tumour DNA was also examined to determine whether the mutation was in the germ line.) The PCR product was purified by DEAE-membrane treatment before DNA sequencing¹³. The sequences of the primer pairs used for PCR corresponded to sequences in the introns flanking each exon; these are given in the legend to Fig. 1. The results are summarized in Table 1. Eight out of sixteen tumour DNA samples showed a single point mutation at codon 249, position 3; seven of these were G→T transversions (arginine→serine) and one was a G→C transversion (arginine→serine). Each mutation was confirmed by sequencing both a second PCR sample and the coding and non-coding strands of the products. No other mutations were found in exons 5, 6 or 8 or the remainder of exon 7. Non-tumour DNA from the eight affected individuals showed a non-mutated genotype. Mutation at codon 249, position 3, has been observed in DNA samples from lung and colon tumours, although the frequency is low^{4,5}. The high mutation frequency (50%) indicates that this position is a hotspot associated with HCC development in this geographic region. The importance of the codon 249 third base pair G→T transversion in HCC may not be a geographically isolated phenomenon¹⁴: of 50% (five out of ten) HCC from South Africa with p53 mutations, three were mutated at codon 249, with the third base pair being a G→T transversion.

TABLE 1 Mutations in p53 in hepatocellular carcinoma

Case	Age	Sex	Grade of HCC	HCC size (cm)	p53 Mutations (codon 249)
84-1	44	M	II	5×4	WT
84-2	33	F	II	14×10	WT
84-3	49	M	III	11×10	AGT
84-4	60	M	II	4×4	WT
84-5	58	F	III	10×7	AGT
89-12	39	F	II	11×8	WT
89-13	35	F	II	5×5	AGC
89-14	65	M	II	11×5	WT
89-15	54	M	II	13×6	AGG/AGT
89-16	37	M	II	10×15	AGG/AGT
90-1	31	M	II	3×3	AGT
90-2	52	M	II	5×4	WT
90-3	35	M	II	6×4	AGT
90-4	39	M	III	5×5	WT
90-5	40	M	III	9×10	WT
90-6	38	M	III	6×7	AGG/AGT

Cases 1–5: The liver DNA samples were isolated in 1984 from patients with primary hepatocellular carcinoma. The rest of the samples were collected from patients in 1989 and 1990. The histological grade was determined²³. WT, Wild type with normal codon 249 (AGG); AGG/AGT, heterozygote with both wild-type and mutant bands present for the third base pair of codon 249; AGT or AGC, mutant sequencing phenotypes giving only one band for the third base pair of codon 249. No mutations were observed in exons 5, 6 and 8 and in the codons of exon 7 other than 249. Non-tumour tissue for all cases in which the tumour was mutated at codon 249 were sequenced for exon 7. All of these samples were wild-type.

Interestingly, aflatoxin B₁ and hepatitis B virus are also risk factors for HCC in South Africa.

Four of our tumour DNA samples with G→T transversions and one with a G→C transversion showed only a single mutant band at position 3 of codon 249 by sequencing. This result most probably represents an effectively homozygous state resulting from the pairing of a mutant p53 allele with a deletion. The sequencing gels for the other three mutated DNA samples showed bands for both the wild-type G and the mutant T at codon 249 position 3. This wild-type G may be derived from contamination of the tumour DNA by adjacent normal liver tissue or tumour stroma. Alternatively, a mutant p53 allele is paired with a second p53 allele which may or may not contain a mutation at some other site. The possibility that this second allele is mutated elsewhere is argued against by the wild type sequencing results we obtained from exons 5, 6, 7 and 8. We found no p53 mutations in any position other than codon 249. These results cannot exclude the possibility that a single p53 mutation at codon 249 position 3, when paired with a normal allele, can be sufficient to express liver tumorigenicity phenotypically.

Cross-contamination of mutated genes, such as from PCR-amplified DNA, in the reaction buffers, primers or genomic DNA preparations, can cause false positive results. To exclude this and to confirm our observations, a new PCR fragment was amplified that had never been used in association with these samples. The upstream primer from exon 6 (5'-CTTATCCGAG-TGGAAGGAAA-3') was paired with the downstream exon 7 sequencing primer to make a 797-base-pair DNA fragment including most of exon 6, all of 7 and the intervening intron. None of the fragments so far described here would have contained sequences complementary to both of these primers. Also, gel purification would separate this fragment from that containing only exon 7. Eight coded samples were analysed in a blind manner. The sequencing results from these PCR products were identical to those shown in Table 1. In addition, six genomic DNA samples from patients 90-1 to 90-6 (Table 1) were isolated and delivered as coded samples for independent amplification and sequencing of exon 7. Results were identical. Also, the DNA

|| To whom correspondence should be addressed.

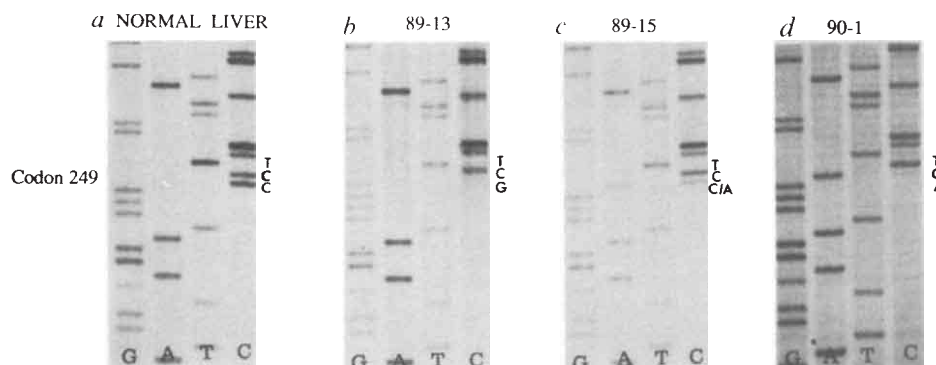


FIG. 1 Sequencing autoradiogram of p53 gene in the region of codon 249 for PCR-amplified DNA samples from normal liver tissue (a) and hepatocellular carcinoma (b, c and d). a, Sequencing of normal liver DNA with normal non-coding codon 249 sequence (AGG). b, Non-coding strand sequences of tumour DNA (89-13) which showed a G→C transversion. c, Non-coding strand sequence of tumour DNA (89-15). Both mutated A band and normal C band at position 3 of codon 249 were observed. d, Non-coding strand sequences of tumour DNA (90-1) which had a G→T transversion.

METHODS. Genomic DNAs were isolated from liver tissues of tumour and normal areas as described previously²⁴. For PCR amplification, genomic DNA (1 µg) was incubated in a total volume of 100 µl of solution containing 50 µM KCl; 10 mM Tris buffer (pH 8.3); 10 mM MgCl₂; 100 pmol oligonucleotide primers and 5 units of *Taq* DNA polymerase (Perkin-Elmer Cetus). Templates were denatured for 5 min at 95 °C, followed by 40 cycles of PCR with incubations of 1 min at 55 °C (annealing), 30 s at 72 °C (polymerization), and 30 s at 94 °C (denaturation). The products of PCR were purified by agarose gel (4%) electrophoresis and DEAE-cellulose paper (S&S, Keen, NH). One-quarter of the purified product was then sequenced by a modification of the dideoxychain-termination method of Sanger *et al.*²⁵, using 1 pmol of

sequencing primers and 5 units of Sequenase. The PCR (pcr) and sequencing (seq.) primers used were as follows:

Exon 5	5'-TGTTCACTTGTGCCCTGACT-3'	left external pcr primer
	5'-TTCAACTCTGTCTCCTTCCT-3'	left internal seq. primer
	5'-AGCAATCAGTGAGGAATCAG-3'	right external pcr primer
	5'-CAGCCCTGTCGTCTCTCCAG-3'	right internal seq. primer
Exon 6	5'-TGGTTGCCAGGGTCCCCAG-3'	left external pcr primer
	5'-GCCTCTGATTCCTCACTGAT-3'	left internal seq. primer
	5'-GGAGGGCCACTGACAACCA-3'	right external pcr primer
	5'-TTAACCCTCCTCCAGAGA-3'	right internal seq. primer
Exon 7	5'-CTTGCCACAGGTCTCCCCAA-3'	left external pcr primer
	5'-AGCGCAGTGGCCTCATCTT-3'	left internal seq. primer
	5'-AGGGTCAAGCGCAAGCAGA-3'	right external pcr primer
	5'-TGTGCAGGTGGCAAGTGGC-3'	right internal seq. primer
Exon 8	5'-TTGGAGTAGATGGAGCCT-3'	left external pcr primer
	5'-TTCTTACTGCCTCTTGCTT-3'	left internal seq. primer
	5'-AGTGTGACTGGAACCTT-3'	right external pcr primer
	5'-AGGCATACTGCACCTTGG-3'	right internal seq. primer

samples isolated from non-tumorous areas of the liver cancer patients whose tumour had a p53 mutation were wild type for codon 249 (AGG). Thus, the high mutation frequency at codon 249 of the p53 gene found in liver tumours of Chinese patients is unlikely to be an artefact.

Our finding that a mutation important in oncogenesis can be confined to a single base pair in a gene and that the change may be almost uniform as to type, was presaged by studies of animals exposed to a specific mutagen. For example, rats exposed to nitrosomethylurea frequently developed mammary carcinomas, of which 20 out of 28 contained a G→A transition in the second base pair of codon 12 of the *H-ras* (15 tumours) of *K-ras* genes (5 tumours)¹⁵. G→A transitions are the only mutations found in mutated *ras* genes in nitrosomethylurea-induced mammary tumours¹⁶.

Hepatitis B virus and aflatoxin B₁ are risk factors for HCC^{17,18} in Qidong¹. Aflatoxin B₁ is a potent chemical carcinogen and mutagen in many animal species¹⁹, which forms deoxyguanosine adducts in rodent and human hepatocytes^{20,21} and frequently causes G→T transversions in mutation assays^{2,3}. Therefore, the

observation of G→T transversions in the third base in codon 249 of p53 in these Chinese HCC has aetiological implications. These can be further explored by examining the mutations in hepatocellular carcinomas induced by aflatoxin B₁ in animal models, including non-human primates, and by exposing normal cultures and SV40 T antigen-'immortalized' human hepatocytes²² to aflatoxin B₁. As all eight of the p53 mutations resulted in the same amino-acid change from arginine to serine, it will also be interesting to transfect an expression vector containing this or other mutant p53 genes into cultured human hepatocytes. Such experiments could help define the pathobiological effects of codon 249 mutant p53 protein, including neoplastic transforming potential and to establish whether the 249 mutant differs in transforming potency from other p53 mutants. Finally, a comparison of p53 mutations in HCC from several geographic areas, for example North America, Europe, Africa and Asia, in which risk factors, including aflatoxin B₁, vary in importance, will be of interest in efforts to link aetiological agents with genetic changes occurring during human hepatocellular carcinogenesis. □

Received 20 December 1990; accepted 21 January 1991.

1. Sun, T., Wu, S., Wu, Y. & Chu, Y. in *Diet, Nutrition and Cancer* (ed. Hayashi, Y.) 225-235 (Japan Sci. Soc. Press, Tokyo, 1986).
2. McMahon, G., Davis, E. F., Huber, L. J., Kim, Y. & Wogan, G. N. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1104-1108 (1990).
3. Sambamurti, K. *et al. Genetics* **120**, 863-873 (1988).
4. Nigro, J. M. *et al. Nature* **342**, 705-708 (1989).
5. Takahashi, T. *et al. Science* **246**, 491-494 (1989).
6. Iggo, R., Gatter, K., Bartek, J., Lane, D. & Harris, A. L. *Lancet* **335**, 675-679 (1990).
7. Hollstein, M., Metcalf, R. A., Welsh, J., Montesano, R. & Harris, C. C. *Proc. natn. Acad. Sci. U.S.A.* **87**, 9958-9961 (1990).
8. Mulligan, L. M., Matlashewski, G., Scrabble, H. J. & Cavenee, W. K. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5863-5867 (1990).
9. Bressan, B. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 1973-1977 (1990).
10. Malkin, D. *et al. Science* **250**, 1233-1238 (1990).
11. Srivastava, S., Zou, Z. Q., Pirolo, K., Buitner, W. A. & Chang, E. H. *Nature* **348**, 747-749 (1990).

12. Soussi, T., Caron de Fromental, C., Mechali, M., May, P. & Kress, M. *Oncogene* **1**, 71-78 (1987).
13. Winberg, G. & Hammarstrand, M. L. *Nucleic Acids Res.* **8**, 253-264 (1980).
14. Bressan, B., Kew, M., Wands, J. & Ozturk, M. *Nature* **350**, 429-431 (1991).
15. Kumar, R., Sukumar, S. & Barbacid, M. *Science* **248**, 1101-1104 (1990).
16. Sukumar, S., Carney, W. P. & Barbacid, M. *Science* **240**, 524-526 (1988).
17. Ganem, D. & Varmus, H. E. *A. Rev. Biochem.* **56**, 651-693 (1987).
18. Harris, C. C. *Cancer Cells* **2**, 146-148 (1990).
19. Wogan, G. N. & Gorelick, N. J. *Envir. Hlth Perspect.* **62**, 5-18 (1985).
20. Wogan, G. N. *Envir. Hlth Perspect.* **81**, 9-17 (1989).
21. Autrup, H. *et al. Chem. Biol. Interact.* **50**, 15-25 (1984).
22. Lechner, J. F., Cole, K. E., Reddel, R. R., Anderson, L. & Harris, C. C. *Cancer Detect. Prev.* **14**, 239-244 (1989).
23. Hsu, I. C. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 2003-2007 (1978).
24. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
25. Edmondson, H. A. & Steiner, P. E. *Cancer* **7**, 462-503 (1954).

Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa

Brigitte Bressac*, Michael Kew†, Jack Wands* & Mehmet Ozturk*‡

* Molecular Hepatology Laboratory, Massachusetts General Hospital Cancer Center, MGH East, 149 13th Street, Charlestown, Massachusetts 02129, USA

† University of The Witwatersrand, York Road, Johannesburg, Parktown 2193, South Africa

HEPATOCELLULAR carcinoma (HCC) is a prevalent cancer in sub-Saharan Africa and eastern Asia¹. Hepatitis B virus and aflatoxins are risk factors for HCC², but the molecular mechanism of human hepatocellular carcinogenesis is largely unknown³. Abnormalities in the structure and expression of the tumour-suppressor gene p53 are frequent in HCC cell lines⁴, and allelic losses from chromosome 17p have been found in HCCs from China⁵ and Japan⁶. Here we report on allelic deletions from chromosome 17p and mutations of the p53 gene found in 50% of primary HCCs from southern Africa. Four of five mutations detected were G→T substitutions, with clustering at codon 249. This mutation specificity could reflect exposure to a specific carcinogen, one candidate being aflatoxin B₁ (ref. 7), a food contaminant in Africa⁸, which is both a mutagen that induces G to T substitution⁹ and a liver-specific carcinogen¹⁰.

Mutations in the p53 gene were investigated in ten primary HCCs from southern Africa. Allelic deletions from chromosome 17p in genomic DNA extracted from tumorous and corresponding non-neoplastic liver tissue were analysed using pYNZ22.1 (ref. 11) and p53 complementary DNA¹² as probes. Tumour-specific loss of one of the 17p alleles was observed in three out of five informative cases (Fig. 1, top; Table 1). Exons 5–8 of the p53 gene, which contain most of the mutations previously identified in diverse human tumours^{13,14}, were amplified by polymerase chain reaction (PCR) from tumour DNA and sequenced. Five tumours had p53 gene mutations: four were G→T substitutions (three at the third base of codon 249 and one at codon 157) and one tumour had a frameshift mutation at codon 286 due to an 8-base pair (5'-GGAAGAG-3') microdeletion (Fig. 1, bottom; Table 1). At least three tumours with p53 mutations had lost the normal p53 allele, as evidenced by our results with the pYNZ22 probe (Fig. 1, top; Table 1). Tumour and non-tumour DNA was analysed after cutting with restriction enzymes; all mutations were found to be somatic and tumour-specific (Fig. 2, top; Table 1).

There was no apparent association of the p53 mutations with the stage of tumour development (early versus late), or with cirrhosis. All the p53 mutations were found in tumours from patients with current or earlier HBV infection (Table 1). Apart from one microdeletion, all mutations were G→T substitutions. The mutation at codon 157 in tumour 6 resulted in a change from valine to phenylalanine. In the remaining three tumours, mutation was at the third base of codon 249, leading to a change from arginine to serine. We also tested six HCC-derived cell lines for the codon 249 mutation by restriction enzyme analysis and nucleic acid sequencing. The G→T mutation at codon 249 was detected in two cell lines (Mahlavu and PLC/PRF/5), both derived from African patients, but not in four others, all derived from non-African patients (Fig. 2, bottom). Although the sample size was limited, our results indicate that p53 mutations in African HCCs are preferentially G→T substitutions with clustering at codon 249.

Other tumour types studied showed a range of p53 gene mutations scattered between codons 132 and 309 (refs 13–15). So far the clustering of p53 mutations at codon 249 seems to be HCC-specific. In support of our observations, G→T mutations at codon 249 have been identified at comparable frequency in HCC from the Qidong area of China¹⁶. Wild-type p53 is a tumour-suppressor gene, but some mutants can behave as dominant oncogenes (reviewed in ref. 17). Thus the high frequency of codon 249 mutations in HCC may indicate that this mutant acts as a dominant oncogene and is therefore selected during the transformation process in hepatocytes. But this is unlikely in view of the fact that we have identified different p53 mutations in two other primary HCCs (namely substitution at codon 157, and a frameshift mutation at codon 286) and found two HCC-derived cell lines (Hep 3B and FOCUS) lacking p53 altogether⁴.

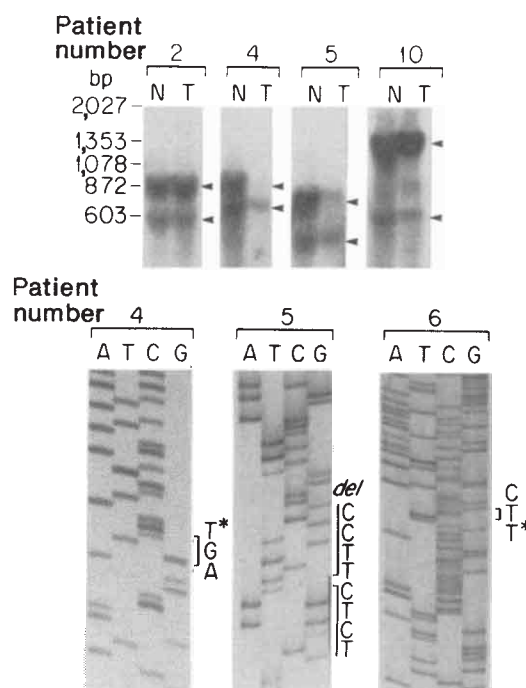


FIG. 1 Top, Allelic analysis of the p53 chromosome region in HCC using pYNZ22.1 as probe¹¹. Non-tumour (N) and tumour (T) DNA from the same patient are shown in adjacent lanes. An obvious change in the ratio of one allele to the other (patients 5 and 10) or the loss of one allele (patient 4) indicates loss of heterozygosity. Total DNA was prepared from frozen liver samples, digested with *Hinf*I, then blotted and probed with ³²P-labelled pYNZ22.1 as described^{11,23}. Bottom, p53 gene mutations in HCCs. The left panel shows the point mutation (T*) at codon 249 detected in tumours of patients 4 (shown here), 9 and 10. The middle and right panels show a microdeletion of 8 bp (del), leading to a frameshift mutation starting at codon 286 (antisense sequence) and a point mutation (T*) at codon 157, respectively. The following sense (F) and antisense (R) series) primers were used for PCR amplification and sequencing analysis of the p53 gene: F1 (5'-ATGCGAATTCCTCCCTGCCCTCAACAAGAT-3'); F3 (5'-GTTGGCTCTGACTGTACCAC-3'); F4N (5'-GCTCAGATAGCGATGGTC-3'); F6 (5'-TGCTCTCCTCCTCTCCTGC-3'); R1 (5'-ATTAGAATTCAGGCGGCTCATAGGGCAC-3'); R2 (5'-TATAGGAATTCGTGGTGAGGCTCCCTT-3'); R3 (5'-CTGGAGTCTTCCAGTGTGAT-3'); R8 (5'-TGAATCTGAGGCATAACTGC-3'). p53 fragments (1.7 kb) were PCR-amplified in duplicate from each tumour DNA as described²³, using F1 and R2 primers. Pooled PCR products were digested with *Eco*RI, phenol-extracted, ethanol-precipitated and purified through a Centricon 30 membrane (Amicon). Fragments were cloned as described²⁴ and sequenced with a T7 sequencing kit (Pharmacia) using F3, F4N and R3, as well as plasmid-derived primers²⁴. To confirm the nature of the mutations, DNAs from tumours 4, 9 and 10 were amplified with F1-R2, purified on Centricon 30 and sequenced directly using primer F3. Likewise, DNAs from tumours 5 and 6 were amplified with F6-R8 and sequenced with F8 (tumour 5) or F6 (tumour 6).

‡ To whom correspondence should be addressed.

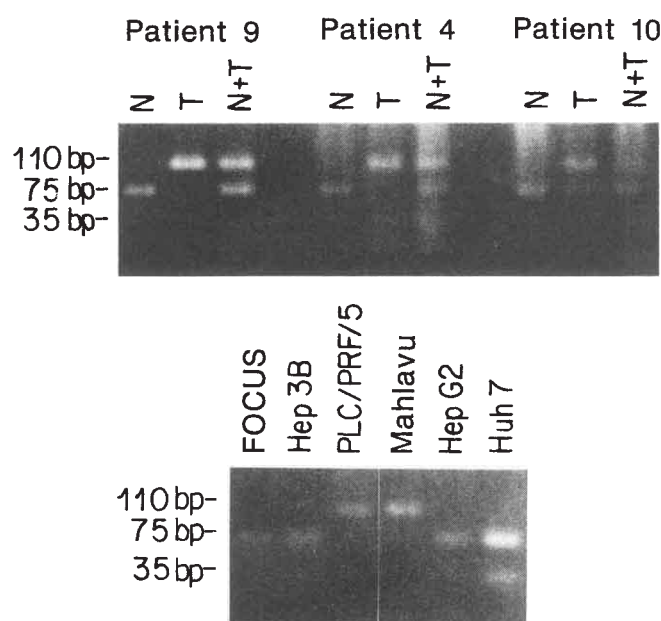


FIG. 2 Top, Restriction enzyme analysis of tumour and non-tumour DNA demonstrates that codon 249 mutations in patients 4, 9 and 10 are somatic. *HaeIII*-digested non-tumour DNAs yield 75- and 35-bp fragments (lanes N). In tumours, mutation 249 leads to the loss of the *HaeIII* site (GGCC → GTCC) and a 110-bp fragment is detected (lane T). Restriction enzyme analysis of a mixture of tumour and non-tumour samples shows that there are no enzyme inhibitors in tumour-derived DNAs (lanes N+T). The same DNAs were 'digestible' with *PleI* (results not shown). Bottom, Restriction enzyme analysis indicating a mutation at codon 249 in two HCC cell lines from African patients (PLC/PRF/5 and Mahlavu)^{25,26}. The mutation in both cell lines was a G → T change at the third base, demonstrated by direct sequencing (not shown). There is no such mutation in cell lines from non-African patients (Focus, Hep 3B, Hep G2 and Huh 7)²⁷⁻²⁹. DNAs were PCR-amplified using F3 and R3 primers, phenol-chloroform extracted, ethanol-precipitated, digested with *HaeIII*, separated on agarose gels, and visualized by ethidium bromide staining.

Mutations of the p53 gene in HCC could also be related to exposure to a specific carcinogen. Two observations are in favour of this hypothesis: first, 4/5 mutations in primary HCCs from southern Africa (our study) and 7/8 mutations in primary HCCs from China¹⁶ are G → T substitutions. Second, two HCC cell lines derived from African patients had G → T mutations at codon 249, whereas four others, all derived from non-African patients, did not. Therefore G → T mutations could be induced by exposure to an environmental factor prevalent in Africa and eastern Asia. Hepatitis B virus and aflatoxins are suspect aetiological factors in HCC development in both of these continents². Integration of hepatitis B virus DNA is a frequent event associated with HCC. A few of these integrations induce insertional mutations in host genes^{18,19} but are unlikely to cause point

mutations. The mutagen aflatoxin B₁ (ref. 7), which is the main food-contaminating aflatoxin species in Africa⁸ and China²⁰, binds preferentially to G residues in G + C-rich regions²¹ and induces G → T substitutions almost exclusively⁹. In this regard, the codon 157 and 249 mutations that we have detected in HCC occurred in G + C-rich regions of the p53 gene and are G → T substitutions. Thus, aflatoxin B₁ could be the causative agent for these p53 G → T mutations. Aflatoxin B₁ is a potent hepatocarcinogen in different species¹⁰ and dietary exposure to it is an epidemiologically defined risk factor for HCC in southern Africa and China², so it is possible that p53 mutations caused by aflatoxins or other environmental carcinogens might contribute to the high incidence of HCC in these areas. □

TABLE 1 Mutations in the p53 gene in hepatocellular carcinomas from southern Africa

Patient	HBV*	Cirrhosis	Tumour†	17p LOH‡	Mutation detected§		
					Codon	Nucleotide	Amino acid
1	No	No	Late	NI			
2	NT	No	Late	No			
3	Yes	No	Early	No			
4	Yes	No	Early	Yes	249	AGG to AGT	Arg to Ser
5	Yes	No	Late	Yes	286	8-bp deletion	Frame shift
6	Yes	No	Late	NT	157	GTC to TTC	Val to Phe
7	Yes	No	Late	NT			
8	Yes	Yes	Early	NI			
9	Yes	Yes	Late	NI	249	AGG to AGT	Arg to Ser
10	Yes	Yes	Late	Yes	249	AGG to AGT	Arg to Ser

* Hepatitis B virus infection as determined by the presence of hepatitis B surface antigen or antibodies against it and/or hepatitis B core antigen in the serum; NT, not tested.

† Tumours were designated as early or late on the basis of tumour size (small, less than 5 cm in diameter; moderately small, 5–10 cm; large, greater than 10 cm) and patient prognosis. Late tumours were large (or massive) and patients died within days or a few weeks after admission to hospital. Early tumours were small or moderately small and surgically resected.

‡ Loss of heterozygosity on chromosome 17p, determined using pYNZ22.1 probe¹¹ (Fig. 1) or by the analysis of p53 gene alleles by southern blotting of *BanII*- or *ScaI*-digested genomic DNA using a human p53 cDNA probe¹² (patient 3). NI, not informative; NT, not tested.

§ Mutations in exons 5 to 8 were determined as described for Fig. 1. To eliminate inconsistencies arising from PCR errors²², control experiments were included. The absence of mutations in tumours 1, 2, 3 7 and 8 was demonstrated by the complete sequencing of 5–7 clones individually. The presence of mutations in tumours 4, 5, 6, 9 and 10 was confirmed by direct sequencing of a second PCR product from each tumour DNA. All mutations were somatic and tumour-specific as determined by restriction enzyme analysis using *HaeIII* for patients 4, 9 and 10 (Fig. 2, top). The tumour-specific loss of the *ThaI* recognition sequence at codon 157 (CGCG → CGCT) for patient 6 was demonstrated by digestion of an F1–R1 primer/PCR product (data not shown; for details of primers, see Fig. 1 legend). Tumour-specific deletion of an 8-bp sequence in patient 5 was demonstrated by analysis of a fragment generated by *HhaI* digestion of an F3–R2 primer/PCR product, which was 45 bp long in the wild-type and 37 bp in the mutant p53 gene (not shown).

Received 27 December 1990; accepted 1 March 1991.

1. Parkin, D. M., Laara, E. & Muir, C. S. *Int. J. Cancer* **41**, 184–197 (1988).
2. Bosch, F. X. & Munoz, N. in *Liver Cell Carcinoma* (eds Bannasch, P., Keppler, D. & Weber, G.) 3–14 (Kluwer Academic, Dordrecht, 1989).
3. Harris, C. C. *Cancer Cells* **2**, 146–148 (1990).
4. Bressac, B. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1973–1977 (1990).
5. Slagle, B. L., Zhou, Y.-Z. & Butel, J. S. *Cancer Res.* **51**, 49–54 (1991).
6. Fujimori, M. et al. *Cancer Res.* **51**, 89–93 (1991).
7. Ames, B. N., Durston, W. E., Yamasaki, E. & Lee, F. D. *Proc. natn. Acad. Sci. U.S.A.* **70**, 2281–2285 (1973).
8. Van Rensburg, S. J. et al. *Brit. J. Cancer* **51**, 713–726 (1985).
9. Foster, P. L., Eisenstadt, E. & Miller, J. H. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2695–2698 (1983).
10. Newberne, P. M. & Butler, W. H. *Cancer Res.* **29**, 236–250 (1969).
11. Nakamura, Y. et al. *Nucleic Acids Res.* **16**, 5707 (1988).
12. Masharani, U., Wolf, D. & Frossard, P. M. *Nucleic Acids Res.* **16**, 7757 (1988).
13. Baker, S. J. et al. *Science* **244**, 217–221 (1989).
14. Nigro, J. M. et al. *Nature* **342**, 705–708 (1989).
15. Chiba, I. et al. *Oncogene* **5**, 1603–1610 (1990).
16. Hsu, I. C. et al. *Nature* **350**, 427–428 (1991).
17. Marshall, C. J. *Cell* **64**, 313–326 (1991).
18. Dejean, A., Bougueleret, L., Grzeschik, K.-H. & Tiollais, P. *Nature* **322**, 70–73 (1986).
19. Wang, J., Chenivresse, X., Henglein, B. & Brechot, C. *Nature* **343**, 555–557 (1990).
20. Yeh, F.-S. et al. *Cancer Res.* **49**, 2506–2509 (1989).
21. Muench, K. F., Misra, R. P. & Humayun, M. Z. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6–10 (1983).
22. Saiki, R. K. et al. *Science* **239**, 487–491 (1988).
23. Malkin, D. et al. *Science* **250**, 1233–1238 (1990).
24. Frohlich, M. et al. *Molec. cell. Biol.* **10**, 3216–3223 (1990).
25. Alexander, J. J., Bey, E., Geddes, E. & Lacetsas, G. S. *Afr. Med. J.* **50**, 2124–2128 (1976).
26. Yang, S. S. et al. *Cancer Res.* **50**, 5658S–5667S (1990).
27. Aden, D. P., Fogel, A., Plotkin, S., Damjanov, I. & Knowles, B. *Nature* **282**, 615–616 (1979).
28. Nakabayashi, H. et al. *Jpn J. Cancer Res./GANN* **75**, 151–158 (1984).
29. He, L. et al. *In Vitro* **20**, 493–503 (1984).

ACKNOWLEDGEMENTS. We thank S. Friend for discussion, and J. Bellamra, S. De la Monte, R. Carlson, K. Galvin and A. Puisieux for their advice and help. This work was supported by grants from the NIH and ARC.

Association of p60^{c-src} with polyoma virus middle-T antigen abrogating mitosis-specific activation

Stefanie Kaech, Lidija Covic, Adrian Wyss & Kurt Ballmer-Hofer*

Friedrich Miescher-Institute, PO Box 2543, CH-4002 Basel, Switzerland

POLYOMA middle-T antigen is required for tumorigenesis in animals and for viral transformation of a variety of cells in culture (reviewed in ref. 1). Middle-T associates with and thereby activates p60^{c-src}, a cellular tyrosine kinase homologous to the oncogene product of Rous sarcoma virus^{2,3}. Activation of p60^{c-src} by middle-T is accompanied both by dephosphorylation of tyrosine 527, a site which negatively regulates src kinase activity (reviewed in refs 4–6) and by autophosphorylation on tyrosine 416 (refs 7–10). Phosphoprotein p60^{c-src} is subject to cell cycle-specific regulation. It is most active during mitosis and repressed in interphase¹¹. Here we report that mitotic p60^{c-src} is dephosphorylated at tyrosine 527. We also show that in cells expressing middle-T, src kinase activity is high both in mitosis and during interphase. An oncogenic mutant src protein, p60^{c-src(527F)}, where tyrosine 527 is substituted by phenylalanine, is also highly active in all phases of the cell cycle.

Earlier reports indicate that activation of p60^{c-src} in mitosis does not change the phosphorylation state of tyrosine 527 and tyrosine 416, sites believed to regulate src kinase activity, and that the transient increase in kinase activity of p60^{c-src} may involve a pathway independent of dephosphorylation of tyrosine 527 (ref. 11). We studied polyoma virus-transformed cells and measured the activity of p60^{c-src} in extracts of synchronized cells using immune-complex kinase assays. Contrary to the transient increase in activity observed during mitosis in normal cells (2.5–9.3-fold), cells expressing middle-T maintained high kinase

activity of p60^{c-src} throughout the cell cycle (Fig. 1b, d and e). This was observed in anti-T or monoclonal antibody 327 immunoprecipitates containing either p60^{c-src} associated with middle-T or the total amount of p60^{c-src}, respectively (Fig. 1d and e). The level of p60^{c-src} expression varied by less than 30% in the various phases of the cell cycle (Fig. 4).

Phosphoprotein p60^{c-src} isolated from mitotic cells had a higher relative molecular mass (M_r), indicative of phosphorylation at several sites in the N terminus that have been implicated as targets of p34^{cdc2} (Figs 1b and 2a; see refs 11–14). *Staphylococcus aureus* V8 protease digestion gave a C-terminal phosphopeptide of M_r 26,000 (26K) in samples phosphorylated *in vitro* (Fig. 3a; see ref. 3). N-terminal phosphopeptides of 18K and 16K and a C-terminal 26K phosphopeptide resulted from digestion of interphase p60^{c-src} labelled *in vivo* (Fig. 3b; see ref. 15). N-terminal peptides derived from mitotic p60^{c-src} showed decreased mobility on SDS-PAGE, probably owing to phosphorylation of threonine 34, threonine 46 and serine 72 as described before^{12,13}. The intensity of the 26K band was reduced in mitotic p60^{c-src}, indicating decreased tyrosine phosphorylation (Fig. 3b).

Immunoprecipitation of p60^{c-src} from C5 cells transformed with middle-T showed that most of the molecules were not associated with middle-T and were autophosphorylated (Fig. 1e, compare panel monoclonal antibody 327 with panel a-T). Furthermore, uncomplexed p60^{c-src} was not activated during mitosis in polyoma-transformed cells (Fig. 1e and g) but was hyperphosphorylated in the N terminus (Figs 1e, 2c and 4c).

The phosphorylation sites of p60^{c-src} were mapped by cyanogen bromide digestion¹⁶. The phosphate content in the N-terminal 31K peptide relative to the C-terminal 4K peptide containing tyrosine 527 was increased from 1.8 ± 0.57 -fold in interphase to 56 ± 33 -fold in mitotic p60^{c-src} (Fig. 3c, lanes I and M). Assuming that p60^{c-src} is phosphorylated at three additional sites in mitosis, we calculated that 50–90 per cent of these kinase molecules were dephosphorylated at tyrosine 527. As suggested by Fig. 1b, we would expect a 5–10-fold activation of p60^{c-src}. The phosphopeptide containing tyrosine 416 (10K) was missing in digests prepared from interphase or mitotic p60^{c-src} (Fig. 3c)¹¹. This suggests that p60^{c-src} is dephosphorylated in mitosis at tyrosine 527 without concomitant phosphorylation at tyrosine 416, implying that these regulatory sites do not depend on each other (Fig. 3c, compare lanes I and M)^{7–10,17–20}.

Mutant p60^{c-src(527F)} isolated from cells in mitosis also had a higher apparent M_r on SDS-PAGE (Figs 1c, 2b, 4d) owing to hyperphosphorylation in the N-terminal domain (Fig. 3b)¹³. But analogous to the middle-T-associated protein, the kinase activity of p60^{c-src(527F)} was not increased during mitosis (Fig. 1c), suggesting that phosphorylation at the putative cdc2 sites is not sufficient to activate this kinase^{12,13}.

We further analysed p60^{c-src} in synchronized cells by western blotting. In cells expressing wild-type c-src protein, p60^{c-src} exists in three forms with slightly different M_r s (Fig. 4b). Most of the material was in the slowest migrating band, whereas in middle-T- and c-src(527F)-expressing cells, mitotic p60^{c-src} was only partially modified with the highest M_r form underrepresented (Fig. 4c–e). This suggests that activation of p60^{c-src} by a mutation mimicking dephosphorylation of tyrosine 527 or complex formation with middle-T influences the phosphorylation state of its cdc2 sites during mitosis. We propose that the N-terminal and C-terminal domains of p60^{c-src} interact with each other, as suggested recently^{6,20}.

We have shown that middle-T association or mutation at tyrosine 527 releases p60^{c-src} from cell-cycle control as these oncogenic forms of the protein cannot be repressed by phosphorylation at tyrosine 527 in interphase. N-terminal phosphorylation or association with middle-T may change the conformation of the protein and cause increased susceptibility to a cellular phosphatase specific for tyrosine 527 or it may prevent phosphorylation by a tyrosine kinase. Alternatively, a tyrosine

* To whom correspondence should be addressed.

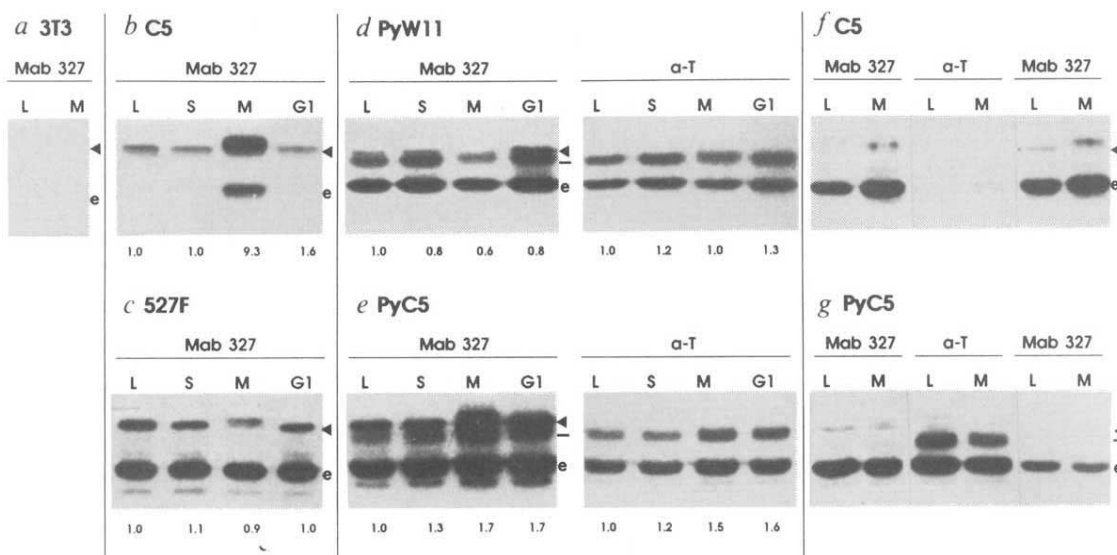


FIG. 1 Kinase activity *in vitro* of p60^{src} immunoprecipitated from synchronized cells. **a**, Control NIH3T3 cells expressing background level of p60^{c-src} activity. **b**, Clone C5, a 3T3 cell line showing normal growth and overexpressing p60^{c-src} about 30-fold²¹. **c**, 3T3 cells transfected with mutant c-src (527F) and expressing p60^{c-src(527F)} about fourfold above the level of the endogenous wild-type src protein. **d**, Middle-T-transformed 3T3 cells, clone W11, overexpressing p60^{c-src} by about fivefold, PyW11. **e**, Middle-T-transformed C5 cells, PyC5. Cell lysates were immunoprecipitated with either monoclonal antibody 327 (Mab 327), specific for p60^{c-src} (ref. 22), or anti-T (a-T) ascites fluid, specific for polyoma virus T antigens. L, cells growing in log phase; S, cells arrested in S phase with 1 mM thymidine for 12 h; M, mitotic cells harvested from cultures arrested with 500 nM nocodazole for 12 h; G1, cells released from mitosis into G1 for 1 h by removal of nocodazole. Synchronization was monitored by flow cytometry and microscopic analysis of mithramycin-stained cells (data not shown). Relative specific kinase activities, normalized for the amount of radioactivity present in enolase in log phase cells, are shown below each lane. The expression of p60^{c-src} did not vary in the cell cycle, as shown in Fig. 4. Activation of p60^{c-src} in mitosis was 4.4 ± 2.8-fold in C5 cells (n = 4), 1.0 ± 0.25-fold in middle-T-transformed cells (n = 6), and 0.8 ± 0.1-fold in c-src(527F)-transfected cells (n = 3). **f**, Controls for the experiment shown in **g**, performed with C5 lysates, demonstrating that mitotic p60^{c-src} is more active than log-phase p60^{c-src} after four rounds of immunoprecipitation. **g**, Separation of middle-T-associated and free p60^{c-src} in PyC5 cells. Log phase (L) and mitotic (M) cell lysates were immunoprecipitated with monoclonal antibody 327 (left panel; assay for free and middle-T-associated p60^{c-src} molecules), a-T (middle panel; assay for middle-T-associated p60^{c-src} only), or three times with a-T followed by monoclonal antibody 327 (right panel; assay for free p60^{c-src}). Kinase activity *in vitro* was assayed as described below.

Arrowheads, position of p60^{c-src}; bars, position of polyoma middle-T antigen; e, position of enolase.

METHODS. The src gene was expressed under the control of simian virus 40 (SV40) early promoter and individual clones were picked from neomycin-resistant (neo^r) mass cultures after transfection into NIH3T3 cells²¹. Middle-T-expressing cells were infected with a Py-MLV retrovirus encoding a middle-T complementary DNA²³. Immunoprecipitation and *in vitro* kinase assays were performed as follows: the cells were washed with phosphate-buffered saline and lysed at 4 °C for 20 min in lysis buffer (120 mM NaCl, 50 mM Tris, pH 8, 1% (w/v) Nonidet P-40, 0.2 trypsin-inhibitor units per ml aprotinin (Sigma), 10 µg per ml leupeptin, 10 µg per ml pepstatin, 0.2 mM EDTA, 50 mM NaF, 200 µM Na₃VO₄, 60 mM NaPPi, 30 mM p-nitrophenyl phosphate, 15 mM β-mercaptoethanol). Lysates were cleared for 10 min at 10,000g, corrected for equal protein content and immunoprecipitated with an excess of antibody for 1 h on ice. In the case of monoclonal antibody 327, a secondary antibody against mouse IgG was added before collecting immunocomplexes with protein A-Sepharose (Pharmacia) for 45 min on ice. Immunoprecipitates were washed twice with TNET (140 mM NaCl, 50 mM Tris, pH 7.5, 5 mM EDTA, 1% (w/v) Nonidet P-40), once with TNE (140 mM NaCl, 50 mM Tris, pH 7.5, 5 mM EDTA) and once with kinase buffer (20 mM Tris, pH 7.5, 5 mM MgCl₂). Kinase reactions *in vitro* were performed with 1 µM [γ-³²P]ATP (100 Ci mmol⁻¹) in 100 µl kinase buffer for 5 min at 30 °C, using 5 µg acid-denatured enolase as an exogenous substrate²⁴. Samples were analysed by 10% SDS-PAGE (minigels, BioRad) after boiling in SDS sample buffer. Gels were exposed on Hyperfilm MP (Amersham) using a Kyokko LH intensifying screen at -70 °C. For the determination of relative specific kinase activities the bands corresponding to enolase were excised and counted in a scintillation counter (Cerenkov). Counts in lanes L were arbitrarily taken as 1.0.

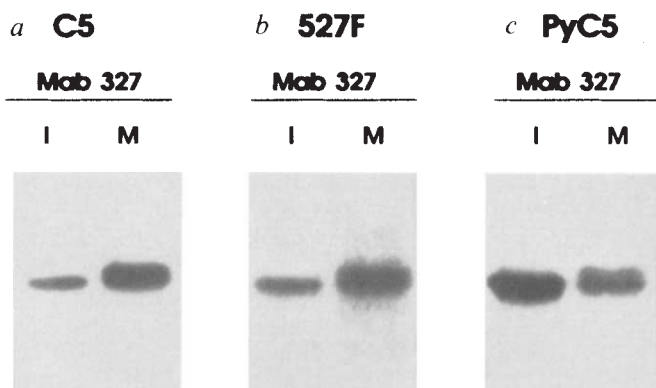


FIG. 2 Immunoprecipitation of *in vivo*-phosphorylated p60^{c-src} from interphase and mitotic cells. **a**, Clone C5, 3T3 cells overexpressing c-src. **b**, c-src(527F)-transfected 3T3 cells. **c**, C5 cells transformed with middle-T, PyC5.

METHODS. Cells arrested in mitosis with 500 nM nocodazole were metabolically labelled for 2 h with 0.5 mCi per ml [³²P]-orthophosphate (Amersham, carrier-free)¹¹. Immunoprecipitates of p60^{src} from lysates of mitotic (M) and interphase (I) cells were analysed by 10% SDS-PAGE as described in Fig. 1. Lysis buffer was further supplemented with 80 mM β-glycerophosphate.

FIG. 3 Peptide mapping of p60^{src}. *Staphylococcus aureus* V8 protease digests of p60^{src} from C5 cells overexpressing c-src and from cells transfected with c-src(527F). *a*, *in vitro*-phosphorylated p60^{src} bands excised from gels shown in Fig. 1*b* and *c*. *b*, *In vivo*-phosphorylated p60^{src}, bands excised from gels shown in Fig. 2*a* and *b*. *c*, Cyanogen bromide map of *in vivo*-phosphorylated p60^{src} from C5 cells, bands isolated from a sample similar to that shown in Fig. 2*a* blotted onto nitrocellulose. *a*–*c*, I, map of p60^{src} from interphase cells, M, map of p60^{src} from mitotic cells; C, map of *in vitro*-phosphorylated p60^{src} from C5 cells showing the tyrosine 416-phosphorylated 10K CNBr peptide. The ratio of the intensity of the 31K to the 4K band was calculated after densitometry of the autoradiograms. This ratio was 1.8 ± 0.57 in interphase and 56 ± 33 in mitotic *src* protein ($n=5$; $\pm$ s.e.m.) The turnover at these phosphorylation sites may vary in the cell cycle and therefore their stoichiometry cannot be determined accurately. *d*, Map of cleavage sites and localization of phosphorylation sites in p60^{src}. METHODS. *In vitro* kinase assays were performed as described in Fig. 1. For analysis of p60^{src} phosphorylated *in vivo*, cells were metabolically labelled with ³²P-orthophosphate as described in Fig. 2. The material was digested with 30 μ g per ml *Staphylococcus aureus* V8 protease after the indicated bands were excised from the gels²⁵. CNBr digestion was performed as described^{16,26} and the fragments were analysed on 16.5% T/6% C separating gels with 10% T/3% C spacer gels according to ref. 27. Densitometric scans of experiments as shown in Fig. 3*c* were performed on a Shimadzu CS 930 scanner.

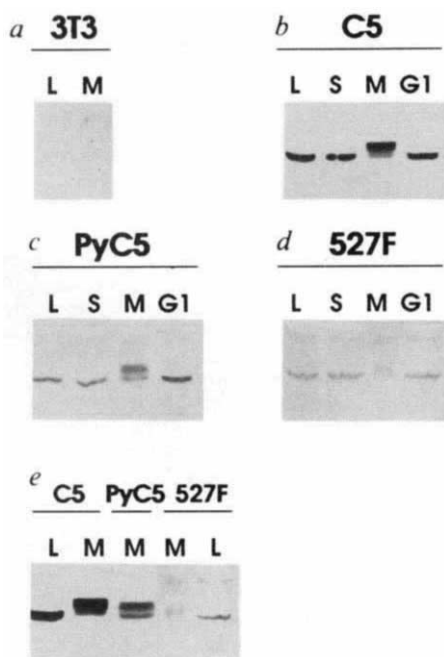
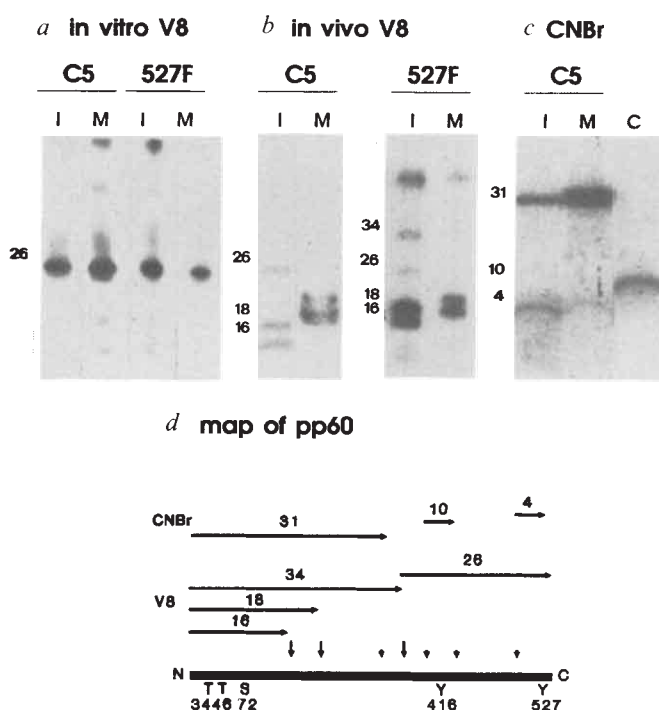


FIG. 4 Western blot analysis of p60^{src} in various phases of the cell cycle. *a*, Control 3T3 cells. *b*, Clone C5, c-src-overexpressing 3T3 cells. *c*, Middle-T-transformed C5 cells, PyC5. *d*, 3T3 cells expressing c-src(527F). *e*, p60^{src} from mitotic cells shown in *b*–*d* shown side by side. Unsynchronized cells (lanes L), cells synchronized in S phase with 1 mM thymidine for 12 h (lanes S) or 500 nM nocodazole in mitosis for 12 h (lanes M), or cells released from mitotic arrest for 30 min (lanes G1) were analysed. METHODS. Synchronization of the cells was verified by flow cytofluorometry (data not shown). The cells were washed, lysed in SDS lysis buffer and the lysates corrected for protein content and analysed by 10% SDS-PAGE. After blotting onto an Immobilon-P membrane (Millipore), p60^{src} was detected with monoclonal antibody 327 followed by alkaline phosphatase-coupled anti-mouse antibody (Southern-Biotech.).

phosphatase activated or transiently expressed in mitosis may dephosphorylate p60^{src}. In this case, modification by p34^{cdc2} might serve a different purpose, for example by affecting the localization of this protein in the cell.

Sustaining increased activity of p60^{src} in interphase and mitotic cells may result in the inappropriate timing of phosphorylation of specific cellular substrates and thus promote uncontrolled cell growth. □

Received 20 August 1990; accepted 5 February 1991.

- Kaplan, D. R., Pallas, D. C., Morgan, W., Schaffhausen, B. & Roberts, T. M. *Biochim. biophys. Acta* **948**, 345–364 (1989).
- Courtneidge, S. A. & Smith, A. E. *Nature* **303**, 435–439 (1983).
- Bolen, J. B. *et al.* *Cell* **38**, 767–777 (1984).
- Parsons, J. T. & Weber, M. J. *Curr. Topics Microbiol. Immun.* **147**, 79–127 (1989).
- Hunter, T. *Cell* **49**, 1–4 (1987).
- Cooper, J. A. in *Peptides and Protein Phosphorylation* (ed. Kemp, B. E.) 86–113 (CRC Press, Cleveland, OH, 1990).
- Cartwright, C. A., Eckhart, W., Simon, S. & Kaplan, P. L. *Cell* **49**, 83–91 (1987).
- Cooper, J. A. & MacAuley, A. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4232–4236 (1988).
- Cooper, J. A., Gould, K. L., Cartwright, C. A. & Hunter, T. *Science* **231**, 1431–1434 (1986).
- Cartwright, C. A., Kaplan, P. L., Cooper, J. A., Hunter, T. & Eckhart, W. *Molec. cell. Biol.* **6**, 1562–1570 (1986).
- Chackalaparampil, I. & Shalloway, D. *Cell* **52**, 801–810 (1988).
- Morgan, D. O., Kaplan, J. M., Bishop, J. M. & Varmus, H. E. *Cell* **57**, 775–786 (1989).
- Sheroy, S. *et al.* *Cell* **57**, 763–774 (1989).
- Norbury, C. J. & Nurse, P. *Biochim. biophys. Acta* **989**, 85–95 (1989).
- Collett, M. S., Erikson, E. & Erikson, R. L. *J. Virol.* **29**, 770–781 (1979).
- Jove, R., Kornbluth, S. & Hanafusa, H. *Cell* **50**, 937–943 (1987).
- Cheng, S. H., Piwnica-Worms, H., Harvey, R. W., Roberts, T. M. & Smith, A. E. *Molec. Cell. Biol.* **8**, 1736–1747 (1988).
- Piwnica-Worms, H., Saunders, K. B., Roberts, T. M., Smith, A. E. & Cheng, S. H. *Cell* **49**, 75–82 (1987).
- Krmieck, T. E. & Shalloway, D. *Cell* **49**, 65–73 (1987).
- Jove, R., Hanafusa, T., Hamaguchi, M. & Hanafusa, H. *Oncogene Res.* **5**, 49–60 (1989).
- Wyss, A., Kaech, S. & Balmer-Hofer, K. *J. Virol.* **64**, 5163–5166 (1990).
- Lipsich, L. A., Lewis, A. J. & Brugge, J. S. *J. Virol.* **48**, 352–360 (1983).
- Donoghue, D. J., Anderson, C., Hunter, T. & Kaplan, P. L. *Nature* **308**, 748–750 (1984).
- Cooper, J. A., Esch, F. S., Taylor, S. S. & Hunter, T. *J. biol. Chem.* **259**, 7835–7841 (1984).
- Cleveland, D. W., Fischer, S. G., Kirschner, M. W. & Laemmli, U. W. *J. biol. Chem.* **252**, 1002–1106 (1977).
- Luo, K., Hurley, T. R. & Sefton, B. M. *Oncogene* **5**, 921–923 (1990).
- Schaeffer, H. & von Jagow, G. *Analyt. Biochem.* **166**, 368–379 (1987).

ACKNOWLEDGEMENTS. We thank J. Brugge for the gift of monoclonal antibody 327, T. Roberts and H. Piwnica-Worms for plasmids carrying the c-src genes and J. Hagmann, T. Schäfer, B. Hemmings, B. Schriener, M. Kamber and R. Ball for critical reading of this manuscript. We also thank B. Heuberger and P. Dollenmeyer for flow-cytofluorometric analysis of synchronized cells, K. v. Hoof for technical assistance, I. Obergföll for photographs and C. Wiedner for secretarial assistance.

A pseudoknot-like structure required for efficient self-cleavage of hepatitis delta virus RNA

Anne T. Perrotta & Michael D. Been*

Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710, USA

* To whom correspondence should be addressed

HEPATITIS delta virus genomic and antigenomic RNAs contain a self-cleavage site hypothesized to function in processing the viral RNA during replication¹⁻³. Self-cleavage requires only a divalent cation¹⁻³ and is mediated at the genomic site by a sequence of less than 85 nucleotides⁴. We propose that the genomic self-cleaving sequence element⁴ and a corresponding sequence from the antigenomic RNA could generate related secondary structures. The region of the antigenomic sequence, predicted from the proposed structure, was synthesized and shown to be sufficient for self-cleavage. Evidence for two stems which form a tertiary interaction was obtained by site-specific mutagenesis of the antigenomic sequence. Efficient self-cleavage in 10 M formamide or 5 M urea, also a property of the genomic sequence⁵, was dependent on base-pairing in both stems. But in the absence of denaturants, the stem distal to the site of cleavage was not required, suggesting that the tertiary interaction stabilizes the structure required for self-cleavage.

The self-cleaving sequences of hepatitis delta virus (HDV) RNA are thought to form secondary structures which differ from previously defined self-cleaving RNA motifs^{1-3,6-8}. The self-cleaving element from the genomic strand RNA of HDV requires only 1 nucleotide 5' to the break site and either 82 nucleotides or, in the presence of denaturants, 84 nucleotides 3' to the break site⁴. Similar potential secondary structures (Fig. 1), consistent with nuclease probing data, can be generated from the sequence

of the genomic element⁴ and the sequence at the antigenomic break site (S. Rosenstein and M.D.B., manuscript in preparation). In both proposed structures, stems I and II together generate a tertiary interaction which is a type of pseudoknot⁹. The proposed structures for the HDV self-cleaving RNAs are distinct from both the hammerhead motif^{6,7} and the single example of a 'hairpin', non-hammerhead, self-cleaving structure⁸.

To test whether the predicted antigenomic sequence was sufficient for self-cleavage, a precursor RNA from position -3 to 87 of the antigenomic sequence was generated by *in vitro* transcription of a cloned synthetic sequence (SA1, Fig. 2a). The purified precursor RNA self-cleaved in 10 mM Mg²⁺, generating a 3' cleavage product that was resolved from the precursor on denaturing polyacrylamide gels (Fig. 3a). The 8-nucleotide-long 5' cleavage product was identified using 5' end-labelled precursor

a SA1 precursor (*Bam*HI run-off)

5' gggaaUUC·GGGUCGGCAU GGCAUCUCCA CCUCCUCGCG³⁰
GUCCGACCUG GGCAUCCGAA GGAGGACG_{Au} CGUCCACUCG⁷⁰
GAUGGCUAAG GGAG⁸⁴ AGCuccggguaccgggggauc

b SA114 precursor (*Bam*HI run-off)

5' ggUC·GGGUC... ...AG GGAG⁸⁴ AGCuccggguaccgggggauc

c SA111 precursor (*Bam*HI run-off)

5' gggC·GGGUC... ...AG GGAG⁸⁴ AGCuccggguaccgggggauc

d SA1-2 precursor (*Hind*III run-off)

5' gggaaUUC·GGGU... ...AG GGAG⁸⁴ caagcu

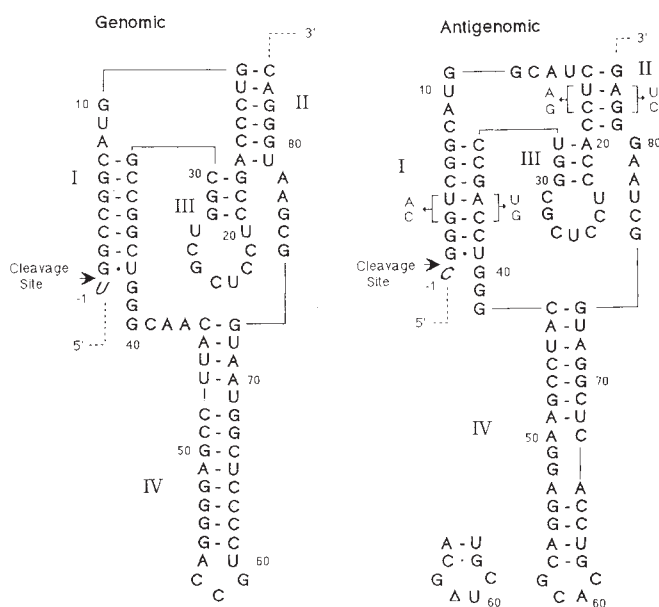


FIG. 1 Potential secondary structures for HDV self-cleaving RNAs. The genomic and antigenomic sequences of HDV are numbered from the break site. Watson-Crick basepairs are indicated by dashes, a G-U pair in each structure by a dot and continuity of sequence by the solid lines. Loop IV of the synthetic antigenomic sequence (Fig. 2) is also shown. Base changes made to test stem I and II in the antigenomic sequence are indicated.

FIG. 2 Synthetic antigenomic sequences. Sequences of transcripts made from *Bam*HI-digested plasmid DNAs a, pSA1; b, pSA114; c, pSA111; and d, *Hind*III-digested pSA1-2. Upper-case letters indicate sequences identical to the HDV antigenomic sequence; lower-case letters indicate sequence derived from the cloning vector (pTZ18U). Two changes to the natural HDV sequence, C59 deleted and A60 changed to U, are also indicated (Fig. 1).

METHODS. Construction of plasmids with a synthetic antigenomic sequence. Four oligonucleotides (ADV-1S, 5'-AATTCGGGTCCGATGGCATCTCCACCT-CCTCGCGTCCGAC; ADV-2S, 5'-CTGGGCATCCGAAGGAGGACGTCGTCCTCCGATGGCTAAGGGAGAGCT; ADV-1A, 5'-TGCCACAGGTCGACCGCAGGAGGT-GGAGATGCCATGCCGACCCG; ADV-2A, 5'-CTCCCTTAGCCATCCGAGTGGACGACGTCCTCCTCGGA) were synthesized on an Applied Biosystems Synthesizer, purified and phosphorylated using T4 polynucleotide kinase and ATP. ADV-1S and 1A (10 pmol each) and ADV-2S and 2A were annealed in separate reactions, mixed and ligated for 12 h at 15 °C using T4 DNA ligase. This reaction was heated to 70 °C for 10 min and then ethanol-precipitated; the DNA was resuspended and digested with restriction endonucleases *Eco*RI and *Sac*I and ligated into pTZ18U, digested with the same two enzymes. The DNA was transformed into *Escherichia coli* (JM83); plasmid DNA from individual isolates was screened by restriction endonuclease digestion and sequenced using dideoxynucleotides and T7 DNA polymerase. Other than the vector-derived sequences, the only other variations from HDV complementary DNA sequences¹⁰⁻¹² were in the region corresponding to the loop at the end of stem IV, where C59 was deleted and A60 converted to a U (Fig. 1). These changes generate a useful unique restriction site in the plasmid DNA at a position of the HDV sequence which is naturally variable. DNA from the resulting plasmid (pSA1) was purified. To generate pSA111 and pSA114, pSA1 DNA was digested with *Eco*RI, treated briefly with mung bean nuclease, religated and transformed into *E. coli* (JM83). To generate pSA1-2, pSA1 DNA was digested with *Sac*I and *Sph*I (removing a *Bam*HI site), treated briefly with T4 DNA polymerase in the presence of all four dNTPs and then ligated. DNA from individual isolates was screened for the loss of either the *Eco*RI site or the *Bam*HI site and then sequenced.

(data not shown). The rate of cleavage of SA1 RNA (*Bam*HI run-off) in 10 mM Mg^{2+} at 37 °C, determined as previously described⁵, was relatively slow (half-life ($t_{1/2}$) = 30 min), but could be enhanced by the addition of denaturant ($t_{1/2}$ = 2 min in 10 M formamide) or by raising the temperature ($t_{1/2}$ = 2 min at 50 °C).

Sequence requirements 5' to the break site were investigated by altering the sequence in that region. Precursors containing only two nucleotides (SA114, Fig. 2b) or one nucleotide (SA111, Fig. 2c) 5' to the cleavage site also cleaved (Fig. 3a), indicating that the identity of the bases 5' to the -1 position was not critical for cleavage activity. The 3' boundary of the sequence required for antigenomic cleavage was defined using 5' end-labelled precursor fragments as previously described^{4,7}. In 12.5 M formamide, the required sequence extended to nucleotide positions 82-84 (the shortest fragments that cleaved). In the absence of denaturant, however, it was two nucleotides shorter, extending to nucleotide positions 80-82 (Fig. 4). As with the genomic sequence⁴, the 3' boundary identified by this method varied slightly with the concentrations of denaturant and Mg^{2+} . Also,

these boundaries are not absolute; a precursor ending at position 79 will cleave, but only in the absence of denaturant and at a substantially reduced rate ($t_{1/2}$ = 2.5 h; data not shown). We have used the 84-nucleotide limit as a minimum for efficient cleavage in 15 min at 37 °C in denaturants, and the plasmid sequence was altered to shorten the 3' end of the transcript (Fig. 2). A *Hind*III run-off from pSA1-2 (Fig. 2d) contains only 6 nucleotides of vector sequence 3' to position 84. At 37 °C in 10 mM Mg^{2+} , this precursor rapidly cleaves in both the absence ($t_{1/2}$ = 16 s) and presence of formamide ($t_{1/2}$ < 10 s) (Fig. 3b, and data not shown).

Base-pairing requirements in the two stems (I and II) that define the tertiary interaction were tested by site-specific mutagenesis of pSA1-2. Three variations of the sequence of stem I were generated. The 5' side of stem I was altered with the combination of a G to C change at position 3 (G3C) and a U to A change at position 4 (U4A); this variant will be referred to as SI5'. A change to the 3' side was made with an A36U change and a C37G change (SI3'). The combination of all four base changes, SI5'3', would restore base-pairing in stem I (Fig. 1). In the absence of denaturants the SI5' and the SI3' precursor RNAs showed reduced levels of cleavage, which were further

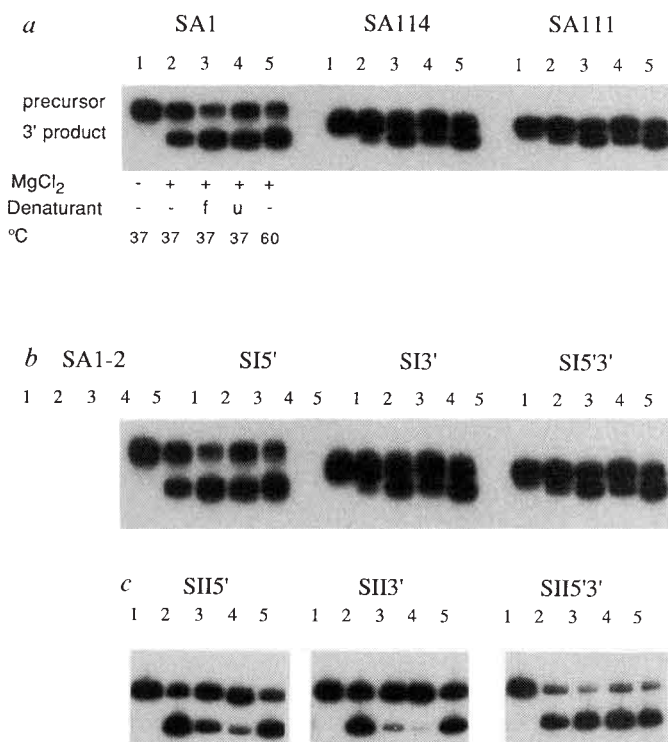


FIG. 3 Self-cleavage of the synthetic antigenomic RNA sequences. **a**, Cleavage of SA1, SA114, and SA111 precursor RNAs. Precursor RNAs (^{32}P -labelled) were incubated in 50 mM Tris-HCl, pH 8.0, 1 mM EDTA only (lanes 1) or with additions of 11 mM $MgCl_2$ (lanes 2 and 5), 11 mM $MgCl_2$ and 10 M (40%) formamide (f) (lanes 3), or 11 mM $MgCl_2$ and 5 M urea (u) (lanes 4). Reactions were incubated for 10 min at 37 °C (lanes 1-4) or 60 °C (lanes 5) and terminated with an equal volume of 95% formamide containing 25 mM EDTA (the reactions are terminated by the chelation of Mg^{2+} , not by the denaturant⁵). Samples were run on a 6% polyacrylamide gel containing 7 M urea⁵. **b** and **c**, Cleavage of SA1-2, stem I variants (SI5', SI3', and SI5'3'), and stem II variants (SII5', SII3' and SII5'3'). Reaction conditions were as described above except the reactions were terminated after 5 min. **METHODS**. Plasmid DNA was digested with *Bam*HI (**a**) or *Hind*III (**b** and **c**), phenol- and chloroform-extracted and ethanol-precipitated. Radiolabelled transcripts were prepared^{4,5} with [α - ^{32}P]CTP included in the transcription reaction. The antigenomic sequence was altered using mutagenic oligonucleotide primers on a uracil-containing single-stranded DNA template^{1,3,14}. The compensatory mutations in stem I were generated in one round of mutagenesis using both oligonucleotide primers. To make the compensatory change in stem II, the original stem II 3' sequence in a SII5' mutant was replaced with a restriction fragment containing the SII3' mutation. Plasmid DNA from each of the variants was purified and sequenced.

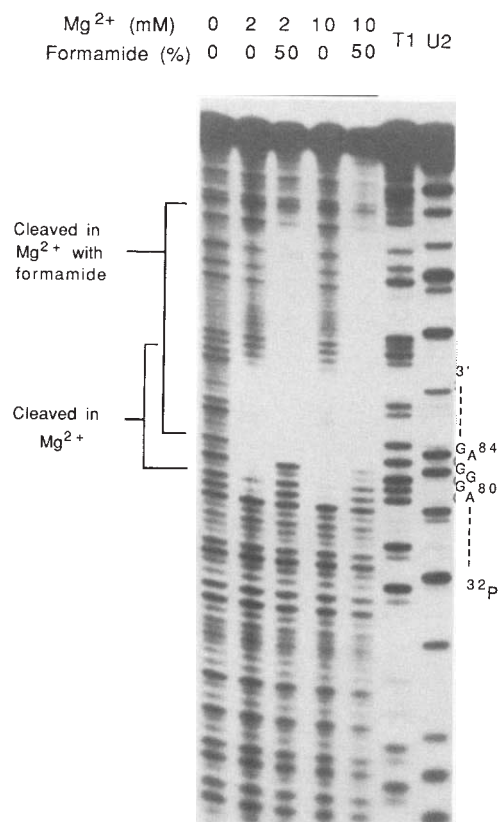


FIG. 4 Determination of the 3' boundary of the self-cleaving antigenomic element. A transcript made from *Hind*III-digested pSA1 DNA was 5' end-labelled with ^{32}P , gel purified and randomly cleaved by partial alkaline hydrolysis to generate a set of labelled fragments with a common 5' end⁴. The *Hind*III run-off contains the same sequence shown in Fig. 2a, but with an additional 30 nucleotides of vector cloning site sequence at the 3' end. This RNA was then incubated at 37 °C for 15 min in 50 mM Tris-HCl, pH 7.5, 1 mM EDTA, or under similar conditions containing in addition: 3 mM $MgCl_2$, 3 mM $MgCl_2$ and 12.5 M formamide, 11 mM $MgCl_2$, or 11 mM $MgCl_2$ and 12.5 M formamide, as indicated. All reactions were terminated by the addition of an equal volume of 95% formamide containing 25 mM EDTA. A G ladder (T1-partial digest) and an A ladder (U2-partial digest) made from the same end-labelled precursor were used as markers. The rationale of the experiment is described by Forster and Symons⁷ and the details of our procedure were as previously described⁴.

reduced with the addition of formamide or urea. However, the S15'3' precursor cleaved to an extent comparable with the SA1 RNA. Therefore, stable base-pairing in stem I was required for efficient self-cleavage activity.

A similar approach was taken for stem II. A S115' mutation was made with a U17A and a C18G change (Fig. 1). This had only a moderate effect on the extent of cleavage in the absence of denaturants, but in the presence of formamide or urea the extent of cleavage was reduced (Fig. 3c). A S113' mutation was made with a G82C and an A83U change. Again, a decrease in the amount of cleavage was observed which was exaggerated by addition of denaturants to the reactions. The S115'3' variant restored the potential for base-pairing in stem II and resulted in similar levels of cleavage to that of the SA1-2 precursor. Cleavage of the S115'3' precursor was not reduced by the addition of denaturants. Therefore, an intact and stable stem II facilitates self-cleavage in the presence of denaturants. This is consistent with the finding that sequences forming the 3' side of that stem are required only in formamide (Fig. 4). We propose that, under favourable conditions, the self-cleaving structure can form in the absence of stem II, but stem II can stabilize the self-cleaving structure. Stem II may therefore contribute to the phenomenon of denaturant-enhanced self-cleavage of certain transcripts⁵ by stabilizing the self-cleaving structure under conditions where competing inactive structures are disrupted. □

Received 12 December 1990; accepted 13 February 1991.

1. Kuo, M. Y.-P., Sharmeen, L., Dinter-Gottlieb, G. & Taylor, J. J. *J. Virol.* **62**, 4439-4444 (1988).
2. Sharmeen, L., Kuo, M. Y.-P., Dinter-Gottlieb, G. & Taylor, J. J. *J. Virol.* **62**, 2674-2679 (1988).
3. Wu, H.-N. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1831-1835 (1989).
4. Perrotta, A. T. & Been, M. D. *Nucleic Acids Res.* **18**, 6821-6827 (1990).
5. Rosenstein, S. P. & Been, M. D. *Biochemistry* **29**, 8011-8016 (1990).
6. Forster, A. C. & Symons, R. H. *Cell* **49**, 211-220 (1987).
7. Forster, A. C. & Symons, R. H. *Cell* **50**, 9-16 (1987).
8. Hampel, A., Tritz, R., Hicks, M. & Cruz, P. *Nucleic Acids Res.* **18**, 299-304 (1990).
9. Pleij, C. W. A. *Trends biochem. Sci.* **15**, 143-147 (1990).
10. Wang, K.-S. et al. *Nature* **323**, 508-514 (1986).
11. Kuo, M. Y.-P. et al. *J. Virol.* **62**, 1855-1861 (1988).
12. Makino, S. et al. *Nature* **329**, 343-346 (1987).
13. Kunkel, T., Roberts, J. D. & Zakour, R. A. *Meth. Enzym.* **154**, 367-382 (1987).
14. Vieira, J. & Messing, J. *Meth. Enzym.* **154**, 3-11 (1987).

ACKNOWLEDGEMENTS. We thank S. Rosenstein and P. Pavco for comments on the manuscript. This work was supported by the NIH. M.D.B. was supported by a Junior Faculty Research Award from the American Cancer Society.

A mediator required for activation of RNA polymerase II transcription *in vitro*

Peter M. Flanagan, Raymond J. Kelleher III,
Michael H. Sayre, Herbert Tschochner
& Roger D. Kornberg

Department of Cell Biology, Fairchild Center,
Stanford University School of Medicine, Stanford, California 94305, USA

ACTIVATOR proteins bind to enhancer DNA elements and stimulate the initiation of transcription. It has been proposed¹⁻³ that activators contact general initiation factors at a promoter, and evidence for such direct interaction has been obtained⁴⁻¹⁰. Studies of transcription *in vitro*, however, have suggested that activators might function through an intermediary molecule(s) distinct from the general factors. In the first of these studies^{11,12}, we exploited the finding that one activator could inhibit transcription stimulated by a second activator (activator interference or 'squenching')¹³⁻¹⁵. This inhibition, which is attributed to competition between the activators for a common target factor, could not be relieved by addition of a large excess of general initiation factors, suggesting that the target for which activators compete is distinct from these

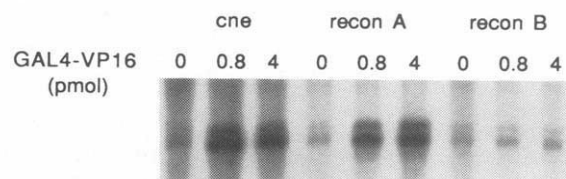
factors. Similar conclusions came from the observation that TFIID's expressed from cloned genes¹⁶⁻¹⁸ fail to replace partially purified 'natural' TFIID fractions in supporting activation, evidently because they lacked some component present in the impure fractions. While these lines of evidence for a novel 'mediator' of activation were negative, we also showed that a partially purified fraction from yeast would reverse activator interference¹². This positive effect of a presumptive mediator provided an assay for its activity, but its role in activation was still only inferred. We now present direct evidence for a mediator which is required for stimulation of transcription *in vitro* by the activators GAL4-VP16 and GCN4, but which has no effect on transcription in the absence of activator protein.

For a compelling demonstration of a mediator between activator proteins and the transcription apparatus, the mediator must be resolved from all components of the transcription apparatus and shown to be required for activation. Efforts to achieve this using a mammalian transcription system have been unsuccessful¹⁷, possibly because of interactions between the mediator and general initiation factors. We now find that mediator activity can be distinguished from that of general initiation factors in a yeast transcription system. Yeast extracts can be resolved into five fractions, designated a, b, c, d and e, all of which are required in addition to RNA polymerase II for initiation at a yeast or mammalian promoter¹⁹. Fraction d can be replaced by cloned, bacterially expressed yeast TFIID (yTFIID). Both the starting extract and the system reconstituted from the five fractions (Fig. 1, cne and recon A, respectively) are responsive to activator proteins. For example, the activator GAL4-VP16, consisting of the DNA-binding domain of yeast GAL4 protein fused to the acidic activation domain of human herpes VP16 (ref. 13), activates transcription 5.3-fold in the extract and 5.7-fold in the reconstituted system (Fig. 1). Further purification of fraction a eliminated the response of the reconstituted system to GAL4-VP16 (Fig. 1, recon B), but did not affect 'basal' transcription in the absence of GAL4-VP16. This suggests that a component required for response to GAL4-VP16 was removed during the purification of fraction a.

The response of the reconstituted system to GAL4-VP16 could be restored by the addition of a DEAE-Sephacel fraction shown previously¹² to relieve activator interference or 'squenching' by GAL4-VP16 (Fig. 2a). The effect was proportional to the amount of the DEAE fraction added within the range tested (Fig. 2b). In contrast, even at the highest levels tested, the DEAE fraction did not appreciably stimulate basal transcription (Fig. 2b), indicating that the fraction does not contain a general stimulatory factor whose effect is unrelated to the presence of activator. This also argues that the DEAE fraction does not merely supply a general initiation factor present in less than saturating amounts in the reconstituted system. Moreover, adding as much as ten times more of fractions a, b, e, yTFIID and RNA polymerase II to the reconstituted reaction had no effect on the level of transcription in the presence of GAL4-VP16 but without the DEAE fraction (Fig. 2c). This shows that all fractions were originally present in saturating amounts in the reconstituted system, and that none contains highly inhibitory material. We conclude that the DEAE fraction supplies a component(s) required for activation by GAL4-VP16 which we refer to as a mediator of activation.

The requirement for a mediator fraction for stimulated transcription in the reconstituted system provides a straightforward assay for mediator activity. Using this assay, the mediator has been enriched from the starting yeast extract about 1,000-fold (Fig. 3, left panel). An amount of purified fraction sufficient for activation contained barely detectable activities of general initiation factors or RNA polymerase (6, 11, 0, 0, 8 and 3% of the amounts of factors a, b, c, yTFIID, e and RNA polymerase II, respectively, required for optimal transcription). As the mediator could be purified in a straightforward manner and the activity did not split in the course of extensive enrichment, it presumably

FIG. 1 Loss of response to the activator GAL4-VP16 during fractionation of a yeast RNA polymerase II transcription system. Transcription reactions contained 100 ng of pGAL4CG⁻, which has a GAL4-binding site upstream of the yeast *CYC1* promoter fused to a G-less sequence²², GAL4-VP16 (ref. 23) in the amounts indicated, and 75 μ g of crude yeast nuclear extract^{12,24} (cne) or the following set of proteins (recon): fractions b (0.02 μ g) and e (6 μ g) from whole-cell extract and fraction c (3 μ g) from nuclear extract, prepared as described¹⁹ with modifications (see below), bacterially expressed yeast TFIID (ref. 25) (yTFIID, 0.4 μ g), and pure yeast RNA polymerase II (ref. 26) (0.04 μ g). Recon A contained fraction a (6 μ g) prepared from whole-cell extract¹⁹; recon B contained Bio-Rex flow-through (1.5 μ g) and 600 mM potassium acetate Bio-Rex eluate (1 μ g) derived from fraction a (see below). Specifically initiated transcripts, processed as described²², formed the major bands in each lane. Radioactivity in specific transcripts (determined with an AMBIS Radioanalytic Imaging System) obtained in the presence of GAL4-VP16 divided by that obtained in its absence (fold stimulation) was: for reaction with cne, 5.3 (0.8 pmol GAL4-VP16) and 3.5 (4 pmol); for recon A, 3.9 (0.8 pmol) and 5.7 (4 pmol); for recon B, 1.0 (0.8 pmol) and 1.0 (4 pmol). Details of fractions a, b and e preparations were as follows. Whole-cell extract was applied to DEAE-cellulose and eluted with 120 and 350 mM ammonium sulphate in buffer A as described¹⁹. The 350 mM eluate was dialysed against buffer A to the conductivity of buffer A containing 100 mM potassium acetate and applied to a Bio-Rex 70 column equilibrated in the same buffer. The column was washed in the same buffer and eluted with buffer A containing 375 and 600 mM potassium acetate.



The flow-through and wash and the 600 mM eluate supplied fraction a activity in recon B. For fraction b, the 120 mM ammonium sulphate eluate from DEAE-cellulose was adjusted to 48% of saturation with ammonium sulphate. The precipitate was suspended in buffer A and passed through a Sephacryl S300 column. Fractions containing activity were pooled and dialysed against buffer A to the conductivity equivalent of buffer A containing 70 mM KCl and applied to a phosphocellulose (P11) column (Whatman) equilibrated in the same buffer. The column was developed with a gradient of 70–700 mM KCl in buffer A, and fraction b activity eluted between 300 and 400 mM KCl. For fraction e, the 120 mM ammonium sulphate eluate from DEAE-cellulose was adjusted to 70% of saturation with ammonium sulphate. The precipitate was suspended in buffer A, dialysed to the conductivity of buffer A containing 100 mM potassium acetate, and applied to a P11 column equilibrated in the same buffer. The column was washed with buffer A containing 300 mM potassium acetate, and fraction e activity was eluted with buffer A containing 700 mM potassium acetate.

comprises one or a small number of components. The possibility that mediator activity resides in a dissociable subunit of one of the general initiation factors or even in a modified form of one of these factors is not excluded.

Two observations indicate that the mediator is a protein, or at least contains an essential protein component. The activity is heat-labile, with a sharply defined transition centred at about 42 °C, and it is abolished by treatment with proteinase K (data not shown).

As the mediator fraction described here is from yeast, and the activation domain of GAL4-VP16 is of human viral origin,

it can be asked whether their interaction is physiologically significant. A yeast activator whose binding site is a thymidine-rich sequence seems to require the same mediator as GAL4-VP16, as its action is subject to interference by GAL4-VP16 (ref. 12). The thymidine-rich-binding activator has not been isolated, however, and little is known about it. GCN4 is a better characterized yeast activator which, like VP16, possesses an acidic activation domain and stimulates transcription in a yeast nuclear extract²⁰. As with GAL4-VP16, stimulation by GCN4 is lost when the extract is fractionated, and restored when the mediator fraction is added back (data not shown).

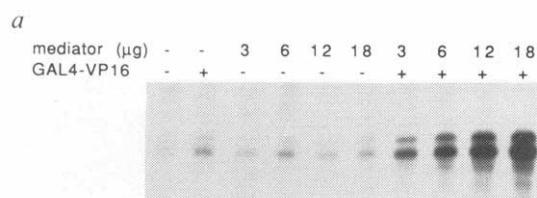


FIG. 2 Restoration of GAL4-VP16 effect on transcription by the addition of mediator. *a*, Transcription reactions as described for recon B in Fig. 1 contained 0.8 pmol of GAL4-VP16 (+) or not (-) and mediator fraction in the amounts indicated. Mediator fraction was the DEAE eluate shown previously to be active in the relief of interference or 'squenching' by GAL4-VP16 (ref. 12). *b*, Radioactivity in specific transcripts in *a* was determined as in Fig. 1 for reactions that contained GAL4-VP16 (activated) or not (basal). *c*, To recon B transcription reactions containing 0.8 pmol of GAL4-VP16 were added extra quantities of transcription factors in multiples of an arbitrary unit of protein defined as follows: fraction a (600 mM Bio-Rex eluate, 1.4 μ g), b (0.6 μ g), e (3 μ g), yTFIID (0.4 μ g), RNA polymerase II (0.04 μ g).

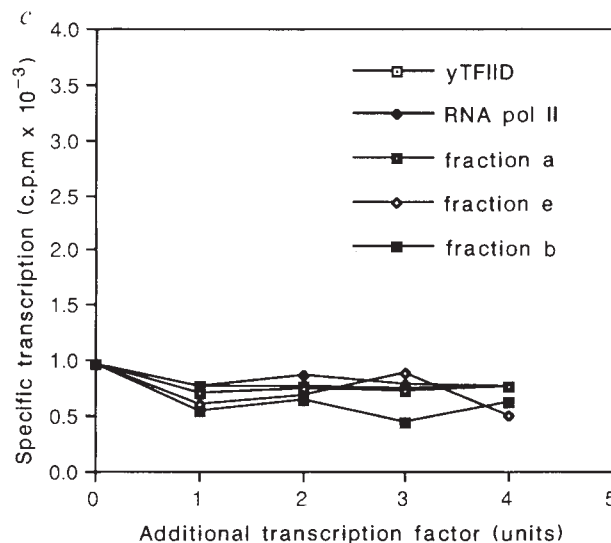
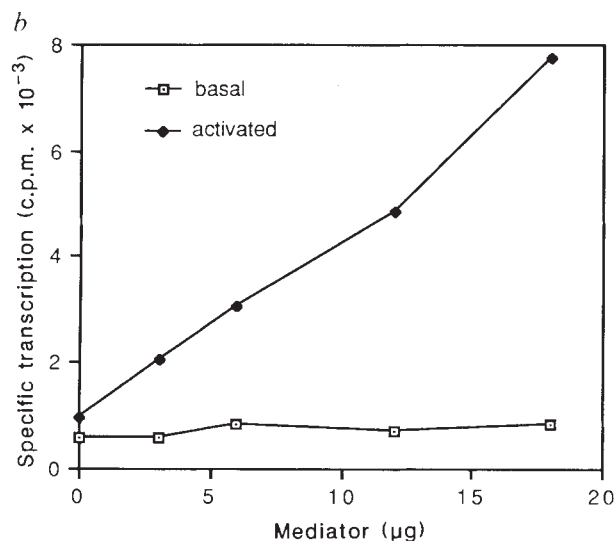
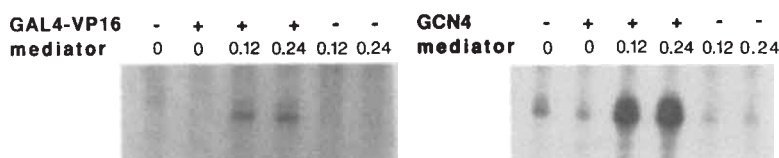


FIG. 3 Partially purified mediator supports activation by both GAL4-VP16 and GCN4. Transcription reactions as described for recon B in Fig. 1 contained (left panel) 100 ng pGAL4CG⁻ and 0.8 pmol GAL4-VP16 or (right panel) 100 ng pS(GCN4)²CG⁻ (this template DNA contains two copies of a consensus GCN4-binding sequence²⁷ upstream of the yeast *CYC1* promoter fused to a G-less cassette in pSP73 (Promega)) and 0.64 pmol homogeneous GCN4. Amounts of mediator fraction added to the various reactions are indicated in μ g. The mediator fraction was derived from yeast whole-cell extract by conventional chromatography on cation and anion exchangers¹⁹, and HPLC on a Bio-Gel SP-5-PW column, as follows. The Bio-Gel SP column was loaded with 25 μ g protein in 0.1 M potassium



acetate and developed with a linear gradient of 0.2–0.6 M potassium acetate in buffer A. The increase in specific activity of the mediator was estimated to be 10-fold, 12-fold, and 9-fold for the three chromatographic steps, giving an overall enrichment of 1,080-fold from the starting extract.

Although the mediator fraction used in these experiments was the relatively crude DEAE fraction described above (Fig. 2, and ref. 12), two observations suggest that GAL4-VP16 and GCN4 use the same mediator. First, mediator activity enriched some 1,000-fold on the basis of activation by GAL4-VP16 also supports activation by GCN4 (Fig. 3, right panel). Second, the two activators interfere with each other (Fig. 4) when increasing amounts of GCN4 (or GAL4-VP16) are added to reactions containing a constant level of GAL4-VP16 (or GCN4) and a template with a single GAL4-binding site (or a template with a pair of GCN4-binding sites). The amount of GCN4 required to cause 50% inhibition of transcription dependent on GAL4-VP16 was ~20-fold greater than the amount of GAL4-VP16 required in the reciprocal experiment, perhaps reflecting a lower affinity of GCN4 for the mediator.

The evidence that a single mediator supports the effects of both GAL4-VP16 and GCN4 raises the possibility that this mediator is required for transcriptional enhancement by all acidic activators. It has been suggested that different mediators are needed by other types of activation domain, such as glutamine- or proline-rich domains¹⁷. Further studies should reveal whether this is the case, and whether or not the classification of activation domains on the basis of amino-acid composition is appropriate. These studies may be facilitated by the recent development of assays for mediators of activation by

GAL4-VP16 (P. Chambon, personal communication) and by USF, Sp1, and NF κ B (R. Roeder, personal communication) in the HeLa transcription system.

We have not demonstrated a direct interaction between the activation domains of GAL4-VP16 and GCN4 and the mediator, and even if these interactions occur, there may be additional direct contacts between activation domains and general initiation factors or RNA polymerase. For example, the data presented here are not incompatible with evidence for direct activator-TFIID interaction^{4–9}.

If mediator is the target of an activation domain, what molecule does the mediator act on to influence transcription? The amino-terminal domains of human and *Drosophila* TFIIDs seem to be required for activation by Sp1 but not for basal level transcription in the HeLa transcription system¹⁸. Similarly, the amino-terminal domain of yeast TFIID is not required for basal transcription in the HeLa system²¹. But a truncated form of yeast TFIID lacking the amino-terminal domain supports activated transcription in the yeast system *in vitro* (R.J.K., P.M.F., D. I. Chasman and R.D.K., manuscript in preparation). The amino-terminal domain of yeast TFIID is thus ruled out as a target, leaving the conserved core of TFIID or other general factors as possible molecules with which the mediator interacts to influence transcription. □

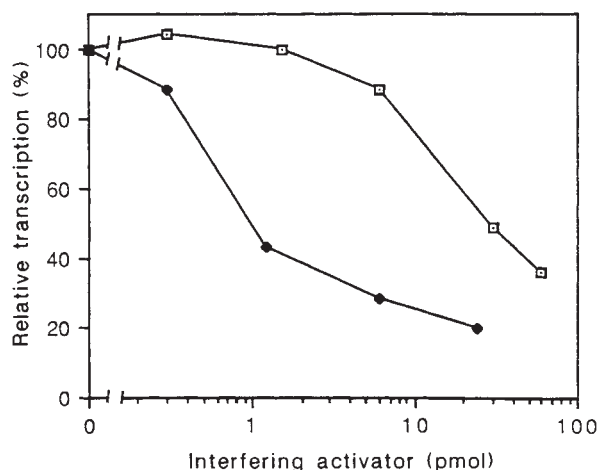


FIG. 4 Activator interference between GAL4-VP16 and GCN4. Transcription reactions with pS(GCN4)²CG⁻ (100 ng) contained an amount of GCN4 protein that caused maximal activation (1.5 pmol), the amounts of GAL4-VP16 (interfering activator) indicated, GAL4-binding oligonucleotide²³ (250 ng, 24 pmol), and cne (64 μ g) (■). Reactions with pGAL4CG⁻ (100 ng) contained an amount of GAL4-VP16 that caused maximal activation (1.2 pmol), the amount of GCN4 (interfering activator) indicated, GCN4-binding oligonucleotide²⁷ (250 ng, 24 pmol), and cne (64 μ g) (□). Radioactivity in specific transcripts, such as for Fig. 1, is expressed as a percentage of that obtained in the absence of interfering activator.

Received 28 November 1990; accepted 5 February 1991.

- Mitchell, P. J. & Tjian, R. *Science* **245**, 371–378 (1989).
- Struhl, K. *Cell* **49**, 295–297 (1987).
- Guarente, L. *Cell* **52**, 303–305 (1988).
- Sawadogo, M. & Roeder, R. G. *Cell* **43**, 165–175 (1985).
- Sawadogo, M. *J. biol. Chem.* **263**, 11994–12001 (1988).
- Horikoshi, M., Carey, M. F., Kakidani, H. & Roeder, R. G. *Cell* **54**, 665–669 (1988).
- Horikoshi, M., Hai, T., Lin, Y.-S., Green, M. R. & Roeder, R. G. *Cell* **54**, 1033–1042 (1988).
- Hai, T., Horikoshi, M., Roeder, R. G. & Green, M. R. *Cell* **54**, 1043–1051 (1988).
- Stringer, K. F., Ingles, C. J. & Greenblatt, J. *Nature* **345**, 783–786 (1990).
- Brandt, C. J. & Struhl, K. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2652–2656 (1989).
- Kornberg, R. D. et al. in *Gene Regulation and AIDS*, (ed. Papas, T. S.) 9–18 (Gulf Publishing, Houston, 1989).
- Kelleher III, R. J., Flanagan, P. M. & Kornberg, R. D. *Cell* **61**, 1209–1215 (1990).
- Trizzenberg, S. J., Kingsbury, R. C. & McKnight, S. L. *Genes Dev.* **2**, 718–729 (1988).
- Gill, G. & Ptashne, M. *Nature* **334**, 721–724 (1988).
- Meyer, M.-E. et al. *Cell* **57**, 433–442 (1989).
- Hoey, T., Dynlacht, B. D., Peterson, M. G., Pugh, B. F. & Tjian, R. *Cell* **61**, 1179–1186 (1990).
- Pugh, B. F. & Tjian, R. *Cell* **61**, 1187–1197 (1990).
- Peterson, M. G., Tanese, N., Pugh, B. F. & Tjian, R. *Science* **248**, 1625–1630 (1990).
- Flanagan, P. M. et al. *J. biol. Chem.* **265**, 11105–11107 (1990).
- Ponticelli, A. S. & Struhl, K. *Molec. cell. Biol.* **10**, 2832–2839 (1990).
- Horikoshi, M., Yamamoto, T., Ohkuma, Y., Weil, P. A. & Roeder, R. G. *Cell* **61**, 1171–1178 (1990).
- Lue, N. F., Flanagan, P. M., Sugimoto, K. & Kornberg, R. D. *Science* **246**, 661–664 (1989).
- Chasman, D. I., Leatherwood, J., Carey, M., Ptashne, M. & Kornberg, R. D. *Molec. cell. Biol.* **9**, 4746–4749 (1989).
- Lue, N. F. & Kornberg, R. D. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8839–8843 (1987).
- Schmidt, M. C., Kao, C. C., Pei, R. & Berk, A. J. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7785–7789 (1989).
- Edwards, A. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2122–2126 (1990).
- Hill, D. E., Hope, I. A., Macke, J. P. & Struhl, K. *Science* **234**, 451–457 (1986).

ACKNOWLEDGEMENTS. We thank C. Ampe and T. Steitz for GCN4 protein, P. Chambon and R. Roeder for communicating unpublished results, J. Feaver, O. Gileadi, and A. Edwards for yeast transcription factors and RNA polymerase, and D. Chasman for GAL4-VP16 and discussions. R.J.K. was supported by the Medical Scientist Training Program from the NIH. M.F.S. was a recipient of an American Cancer Society postdoctoral fellowship. H.T. was a recipient of a Deutsche Forschungsgemeinschaft fellowship. This research was supported by the NIH.

Blocking of Tat-dependent HIV-1 RNA modification by an inhibitor of RNA polymerase II processivity

Martin Braddock*, Andrew M. Thorburn*†, Alan J. Kingsman*‡§ & Susan M. Kingsman*

* Virus Molecular Biology Group, Department of Biochemistry, South Parks Road, Oxford OX1 3QU, UK

‡ British Biotechnology Ltd, Watlington Road, Cowley, Oxford OX4 5LY, UK

HUMAN immunodeficiency virus gene expression is regulated transcriptionally and post-transcriptionally¹⁻⁴ by the virally encoded *tat* protein (Tat). Tat functions through an RNA target sequence located in the untranslated region at the 5' end of viral transcripts⁵⁻⁸. In *Xenopus* oocytes, translation of RNA containing the target sequence is specifically activated by Tat. This activation only occurs if the RNA is injected into the nucleus, and might be due to a Tat-dependent, nucleus-specific chemical modification of the RNA which somehow facilitates translation⁸. Here we demonstrate that Tat activation of its target RNA in the nucleus involves a Tat-dependent covalent modification. The modified RNA is competent for translation after reinjection into either the nucleus or the cytoplasm in the absence of Tat. Furthermore, we find that the nucleoside analogue 5,6-dichloro-1- β -D-ribofuranosylbenzimidazole, which inhibits processivity of RNA polymerase II (ref. 9), blocks this Tat-dependent modification.

We synthesized an RNA (TAR-CAT RNA) with the Tat target sequence fused to the bacterial chloramphenicol acetyltransferase coding sequence⁸ (Fig. 1a). When this TAR-CAT RNA was injected into the nuclei or cytoplasm of *Xenopus* oocytes it was not translated in the absence of Tat (Fig. 2a, lanes 2 and 4). If Tat was present, the RNA was still not translated when injected into the cytoplasm (Fig. 2a, lane 3) but it was when injected into the nucleus (Fig. 2a, lane 1). Therefore the translation of TAR-CAT RNA needs Tat and a nucleus-specific event. To test whether TAR-CAT RNA is altered in the presence of Tat after injection into the nucleus, we rescued the RNA, now called TNI-RNA (Fig. 1b), and reinjected it into the cytoplasm or nucleus of fresh oocytes. Our experimental strategy and nomenclature is shown in Fig. 1b. If there was no alteration no CAT enzyme would be produced from these reinjections. The result would be the same as after the initial injections without Tat (Fig. 2a, lanes 2 and 4); we obtained the opposite result, in that the rescued RNA was efficiently translated (Fig. 2a, lanes 5 and 9). These data suggested that the TAR-CAT RNA had been altered as a result of passage through the nucleus in the presence of Tat. To confirm that this effect was Tat-specific, RNA was rescued from oocytes after first injecting TAR-CAT RNA into the nucleus without Tat. This rescued RNA, called NI-RNA, was not translated when reinjected into the cytoplasm (Fig. 2a, lane 6) or the nucleus (Fig. 2a, lane 8). The rescued RNAs were recovered at comparable levels and were unchanged at the 5' ends (Fig. 2b), suggesting that the differences were not due to differential recovery or degradation. We also tested for the functional integrity of the NI-RNA by reinjecting it into the nucleus and supplying Tat. The NI-RNA was now translated (Fig. 2a, lane 7), indicating that it was not inherently untranslatable but, as with the original input TAR-CAT RNA (Fig. 2a, lanes 1-4), the translation was dependent on Tat and the nucleus. To confirm that passage through the nucleus was necessary, TAR-CAT RNA was injected into the cytoplasm in the presence and absence of Tat and then the rescued RNA, designated respec-

tively TCI-RNA and CI-RNA, was reinjected into oocytes. Neither of the rescued RNAs was translated when injected into the cytoplasm (Fig. 2a, lanes 10 and 13) or the nucleus (Fig. 2a, lanes 11 and 14), but both were translated when injected into the nucleus in the presence of Tat (Fig. 2a, lanes 12 and 15), again indicating that the RNA was intact. Passage of RNA through the cytoplasm in the presence or absence of Tat therefore had no effect on the subsequent translation of the RNA. We conclude that the normally translationally repressed TAR-CAT RNA can be converted to a translatable form by passage through the nucleus in the presence of TAT. The rescued RNA was treated with protease, DNase and denaturants, so increased translation must reflect a direct covalent modification of the RNA that is mediated by Tat and a nucleus-specific factor. Modifications to RNA that could affect translation include alterations to the 5' cap structure¹⁰ and enzymatic modification of bases, such as occurs in transfer RNA¹¹ and during the unwinding of some RNA structures¹²⁻¹⁴.

We have previously shown that in *Xenopus* oocytes Tat activation of a TAR-CAT plasmid carrying the human immunodeficiency virus (HIV) long terminal repeat (LTR) is resistant to the transcription inhibitors actinomycin D, cordycepin and α -amanitin, provided that RNA accumulates before addition of inhibitor. This indicated that Tat functions largely

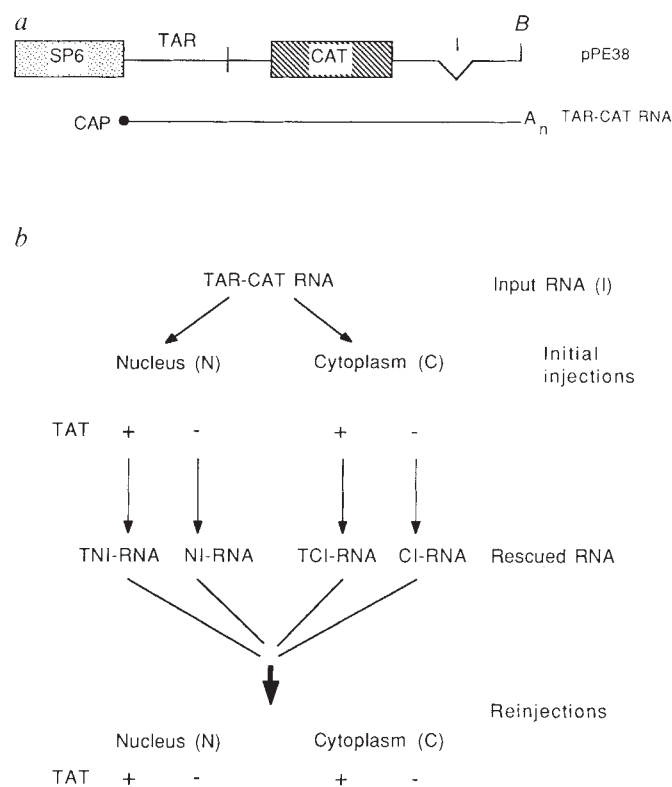


FIG. 1 Plasmid constructs and experimental design. *a*, Plasmid pPE38 has been described previously^{8,19}. It contains the SP6⁸ transcription cassette shown schematically. Transcription with SP6 polymerase, *in vitro*, produces an RNA that has the authentic TAR sequence from nucleotides +1 to +82 fused to the bacterial chloramphenicol acetyltransferase (CAT) coding sequence⁸. *B*, BamHI run-off site; I, SV40 small *t* intron. The presence of an intron does not influence the translation of this RNA⁸. *b*, Outline of the experimental scheme. TNI, RNA that experienced Tat (T) and the nucleus (N) at the initial injection (I); NI, RNA that experienced the nucleus at the initial injection; TCI, RNA that experienced Tat and the cytoplasm (C) at the initial injection; CI, RNA that experienced the cytoplasm at the initial injection; I, the *in vitro* synthesized input RNA.

METHODS. *a*, TAR-CAT RNA was produced by *in vitro* transcription with SP6 polymerase and then capped and polyadenylated (A_n) as described previously^{8,19}.

§ To whom correspondence should be addressed.

† Present address: Basic Science Building, Room 5084, School of Medicine, University of California San Diego, La Jolla, California 92093-0636, USA.

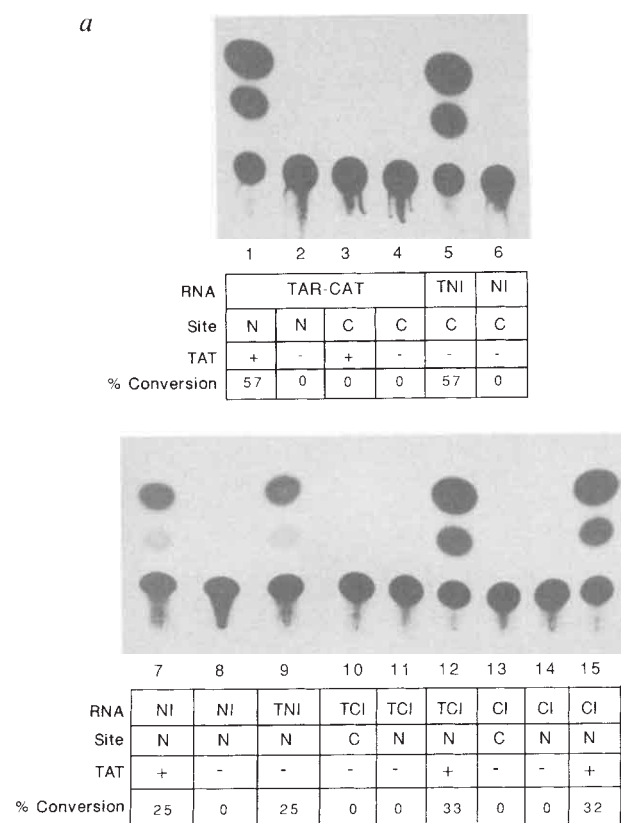


FIG. 2 Translation of TAR-CAT RNA and rescued RNAs. *a*, CAT assays. *b*, Primer extension RNA analysis. METHODS. *a*, Stage-6 *Xenopus* oocytes were prepared as described⁸. RNA (3 ng) was injected into the nucleus or cytoplasm. Titration shows that this

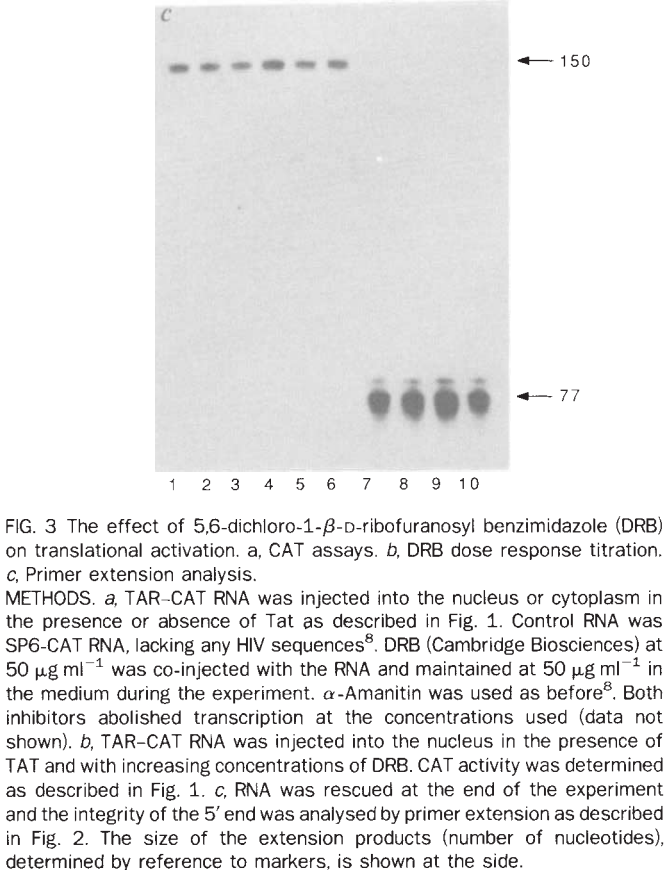
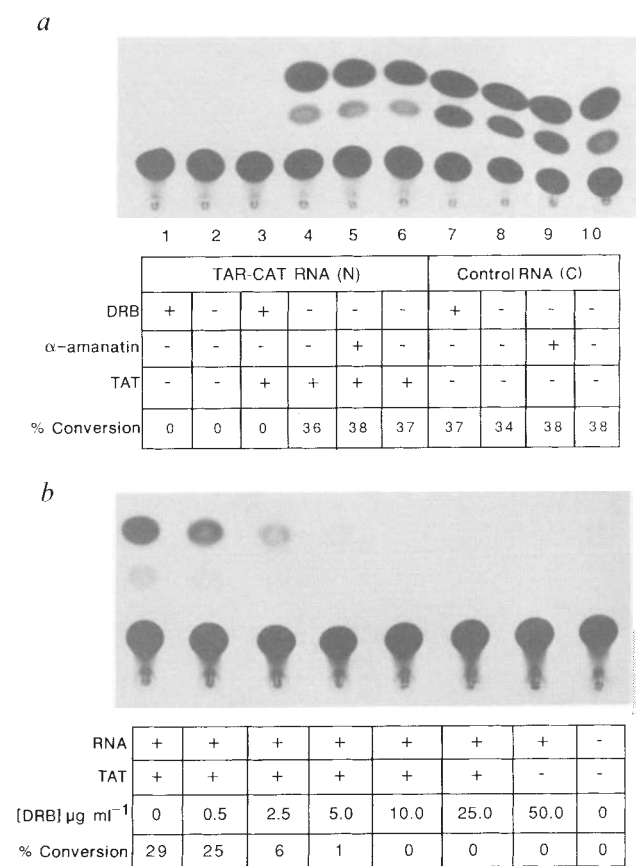
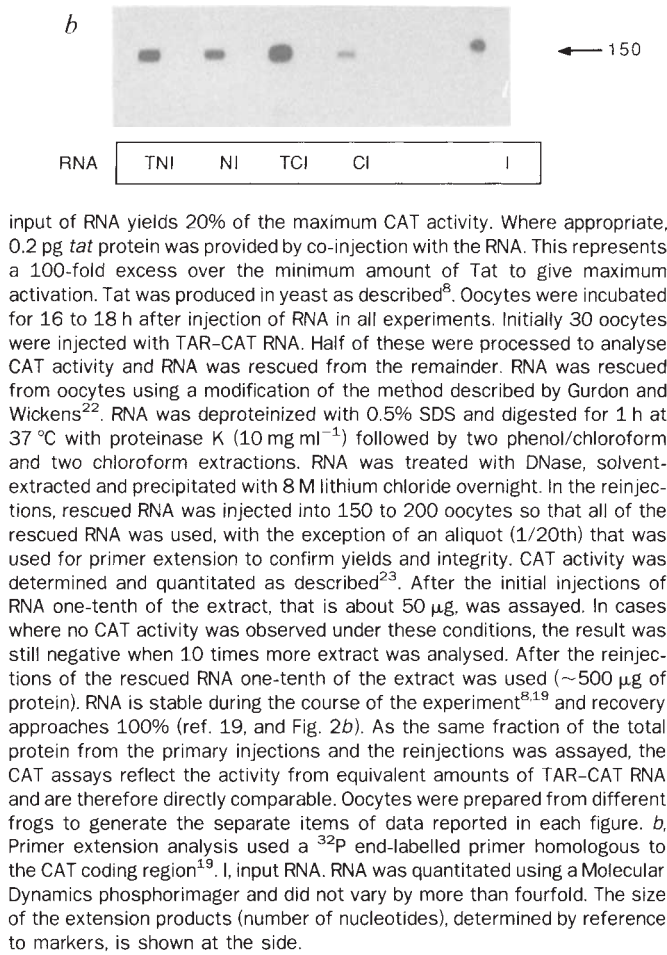


FIG. 3 The effect of 5,6-dichloro-1-β-D-ribofuranosyl benzimidazole (DRB) on translational activation. *a*, CAT assays. *b*, DRB dose response titration. *c*, Primer extension analysis. METHODS. *a*, TAR-CAT RNA was injected into the nucleus or cytoplasm in the presence or absence of Tat as described in Fig. 1. Control RNA was SP6-CAT RNA, lacking any HIV sequences⁸. DRB (Cambridge Biosciences) at 50 µg ml⁻¹ was co-injected with the RNA and maintained at 50 µg ml⁻¹ in the medium during the experiment. α-Amanitin was used as before⁸. Both inhibitors abolished transcription at the concentrations used (data not shown). *b*, TAR-CAT RNA was injected into the nucleus in the presence of TAT and with increasing concentrations of DRB. CAT activity was determined as described in Fig. 1. *c*, RNA was rescued at the end of the experiment and the integrity of the 5' end was analysed by primer extension as described in Fig. 2. The size of the extension products (number of nucleotides), determined by reference to markers, is shown at the side.

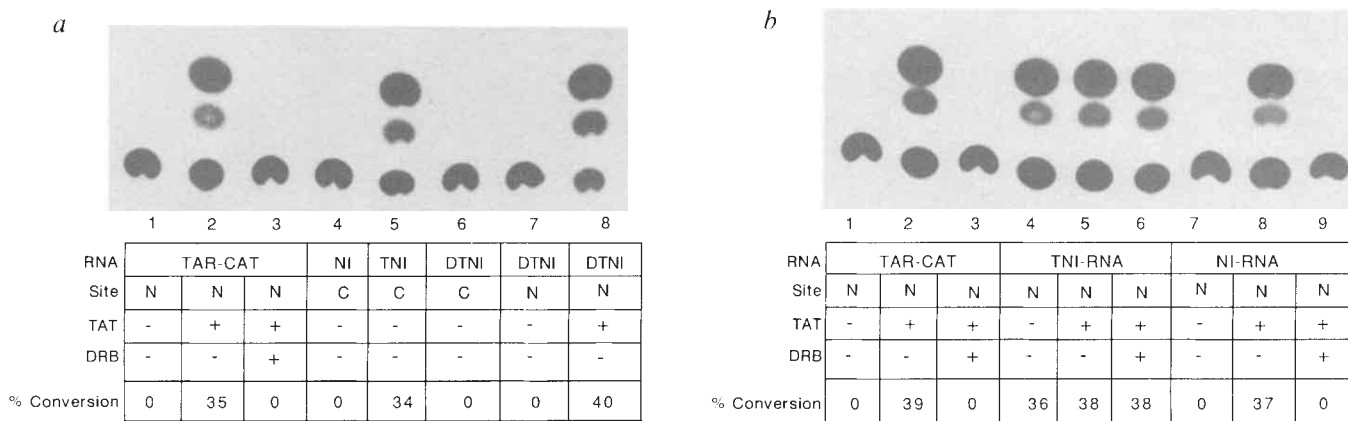


FIG. 4 DRB-mediated inhibition of Tat-dependent RNA modification. **a**, CAT assays to show the translation of TAR-CAT RNAs rescued from oocytes that had been previously injected into the nucleus with TAR-CAT RNA in the presence of Tat (TNI-RNA), in the presence of Tat and DRB (DTNI-RNA) and in the absence of Tat (NI-RNA). Lanes 1–3 reproduce the primary observation that DRB inhibits Tat activation of TAR-CAT RNA. **b**, CAT assays

to show the effect of DRB on the translation of modified (TNI-RNA) and unmodified (NI-RNA). Lanes 1–3 reproduce the primary observation that DRB inhibits Tat activation of TAR-CAT RNA.

METHODS. TAR-CAT RNA and rescued RNAs were injected into oocytes and analysed as described in Fig. 1. DRB was used as described in Fig. 3.

post-transcriptionally in this system⁸. We have also found that the transcription inhibitor 5,6-dichloro-1- β -D-ribofuranosylbenzimidazole (DRB) blocks Tat activation (unpublished data). Taken together, these results suggest that DRB is able to inhibit a post-transcriptional event. We therefore investigated whether DRB might affect RNA modification. First, to confirm that DRB functions post-transcriptionally, TAR-CAT RNA was injected into the nucleus of *Xenopus* oocytes with and without Tat and with and without DRB (Fig. 3). DRB-blocked Tat-mediated translational activation of TAR-RNA (Fig. 3a, compare lanes 3 and 4) but had no effect on general translation as indicated by the efficient expression of a control CAT RNA (Fig. 3a, lanes 7–10). The 50% inhibitory concentration (IC_{50}) was about $2 \mu\text{g ml}^{-1}$ (Fig. 3b). In contrast, α -amanitin had no effect on Tat activation of translation of TAR-CAT RNA (Fig. 3a, lane 5), indicating that a general transcription block did not cause the effect. RNAs rescued after DRB treatment were the same as input RNA, as judged by primer extension analysis (Fig. 3c).

To test whether DRB inhibited TAR-RNA modification, TAR-CAT RNA was injected into the nucleus in the presence of TAT and DRB. The rescued RNA (DTNI-RNA) was re-injected into the cytoplasm (Fig. 4a, lane 6): there was no translation, but in the same experiment TNI-RNA was fully translated (Fig. 4a, lane 5). DTNI-RNA was not translated when re-injected into the nucleus without Tat (Fig. 4a, lane 7) but after addition of Tat it was fully translated (Fig. 4a, lane 8). This latter result shows that DRB does not simply inactivate the RNA, for example by degradation. The effect of DRB on the translation of RNA that had already been modified, TNI-RNA, was also tested. As before, this RNA is translated equally efficiently with or without Tat (Fig. 4b, lanes 4 and 5) and translation is unaffected by DRB (Fig. 4b, lane 6). The modified RNA was therefore resistant to DRB. As expected, the transla-

tion of NI-RNA (that is, unmodified) was not activated by Tat in the presence of DRB (Fig. 4b; compare lanes 8 and 9), although it was fully activated by Tat in the absence of DRB (Fig. 4b, lane 8). The translational activation of TAR-CAT RNA by Tat in *Xenopus* oocytes therefore involves covalent modification of the RNA and this modification involves a DRB-sensitive step.

The ability of Tat to activate the two apparently spatially and temporally distinct processes of transcription^{15–18} and translation^{2,5,8} is an enigma. Our previous work suggested that the translational fate of an RNA can be linked to transcription because the requirement for Tat activation of TAR-RNA depends on upstream promoter sequences¹⁹. If there is a link between translational fate of an RNA and the quality of its transcription, then Tat could affect both through a common step. The idea of a common step is consistent with the findings that RNA is the target for both transcriptional^{18,20} and translational activation⁸ by TAT, that the nuclear machinery is required for translational⁸ and transcriptional activation, and that the transcriptional inhibitor DRB also functions to block the translational activation. Furthermore, Tat seems to activate transcription by increasing polymerase processivity²¹, and DRB interferes with processivity⁹. A DRB-sensitive step in transcription and in translational activation could be central to a mechanism of Tat activation with effects on both transcription and translation. Identification of the cellular factors that are inhibited by DRB may therefore lead to the development of a model for Tat action that unifies the disparate observations. The biochemical characteristics of TAR modification and the mechanism of recruitment of the modifier by Tat will provide insight into translational control in general and Tat activation in particular. This may facilitate the rational design of antiviral drugs to inhibit Tat action. □

Received 7 December 1990; accepted 13 February 1991.

- Dayton, A. I., Sodroski, J. G., Rosen, C. A., Goh, W. C. & Haseltine, W. A. *Cell* **44**, 941–947 (1986).
- Feinberg, M. B., Jarrett, R. F., Aldovini, A., Gallo, R. C. & Wong-Staal, F. *Cell* **46**, 807–817 (1986).
- Rosen, C. A. *et al.* *Nature* **319**, 555–559 (1986).
- Wright, C. M., Felber, B. K., Paskalis, H. & Pavlakis, G. N. *Science* **234**, 988–992 (1986).
- Rosen, C. A., Sodroski, J. G. & Haseltine, W. A. *Cell* **41**, 813–823 (1985).
- Hauber, J. & Cullen, B. R. *J. Virol.* **62**, 673–679 (1988).
- Muesing, M. A., Smith, D. H. & Capon, D. J. *Cell* **48**, 691–701 (1987).
- Braddock, M. *et al.* *Cell* **58**, 269–279 (1989).
- Chodosh, L. A., Fire, A., Samuels, M. & Sharp, P. A. *J. Biol. Chem.* **264**, 2250–2257 (1989).
- Darzynkiewicz, E. *et al.* *Nucleic Acids Res.* **16**, 8953–8962 (1988).
- Elliott, M. S. & Trewyn, R. W. *J. Biol. Chem.* **259**, 2407–2410 (1984).
- Bass, B. L. & Weintraub, H. *Cell* **55**, 1089–1098 (1988).

- Wagner, R. W., Smith, J. E., Cooperman, B. S. & Nishikura, K. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2647–2651 (1989).
- Kimmelman, M. & Kirschner, M. W. *Cell* **59**, 687–696 (1989).
- Cullen, B. R. *Cell* **46**, 973–982 (1986).
- Laspi, M. F., Rice, A. P. & Mathews, M. B. *Cell* **59**, 283–292 (1989).
- Berkhout, B., Silverman, R. H. & Jaeng, K.-T. *Cell* **59**, 273–282 (1989).
- Selby, M. J. & Peterlin, B. J. *Cell* **62**, 769–776 (1990).
- Braddock, M. *et al.* *Cell* **62**, 1123–1133 (1990).
- Southgate, C., Zapp, M. L. & Green, M. R. *Nature* **345**, 640–642 (1990).
- Marciniak, R. A., Calnan, B. J., Frankel, A. D. & Sharp, P. A. *Cell* **63**, 791–802 (1990).
- Gordon, J. P. & Wickens, M. P. *Meth. Enzym.* **101**, 370–386 (1983).
- Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044–1051 (1982).

ACKNOWLEDGEMENTS. This research was supported by an SERC/DTI LINK grant with British Biotechnology Ltd and Glaxo Group Research.

Oligonucleotide-based therapeutics

Michael L. Riordan and John C. Martin

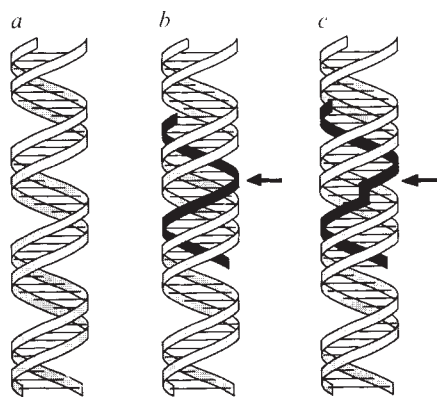
Oligonucleotide analogues offer several mechanisms for potential therapeutic action; triple helix agents that inhibit disease-causing genes at the DNA level show particular promise.

TECHNICAL advances over the past decade in the synthesis and molecular biology of oligonucleotide analogues have spurred a research effort dedicated to the discovery of oligonucleotide drugs. Three classes of potential therapeutic agents have been identified. The first two classes, antisense oligonucleotides and triple helix agents, are known as "code blockers" because their mechanism of therapeutic action involves blocking the expression of the genetic code. The third class of oligonucleotides with therapeutic potential has been termed "aptamers"¹. These are selected and amplified oligonucleotides that have been isolated from random pools of synthetic oligonucleotides according to their ability to bind with high affinity to biological target molecules.

Code blockers

Antisense oligonucleotide analogues are designed to bind to RNA by Watson-Crick hybridization and block RNA processing or translation, and thus the synthesis of a disease-mediating protein. The considerable drug discovery effort devoted to this approach followed the demonstration that antisense oligonucleotides could down-regulate protein synthesis in cells^{2,3} and has been the subject of a recent excellent review⁴. Calculations and empirical sequence searches have shown that an oligonucleotide needs to be 15–18 nucleotides in length to specify a unique target, thus requiring a molecular weight of the order of 5,000. These substances have been extensively modified to improve their pharmaceutical properties. Derivatizations of the 3' terminus of oligonucleotides have been found to improve greatly serum stability to exonucleases⁵. Replacement of the phosphate group by the isosteric formacetal group produces analogues with an achiral backbone, which is resistant to nucleases and yet still able to hybridize to RNA⁶. The formacetal group is also a neutral replacement for the charged phosphate moiety and is likely to enhance cellular permeation. However, efficient permeation remains the greatest challenge for oligonucleotide drug development. Oligonucleotides that freely permeate cells are needed to allow for the low effective doses that will make antisense therapy economically feasible.

Less well known than antisense is the related approach of blocking gene expression



a, Schematic structure of duplex DNA; b, an oligonucleotide (dark strand) binding to duplex DNA to form a triple helix; and c, a switchback oligonucleotide (dark strand) interacting with both strands of duplex DNA.

by inhibition of replication or transcription of DNA with triple helix-forming oligonucleotide analogues (see figure). These analogues bind in a sequence-specific manner in the major groove of the duplex structure of DNA. For the most widely studied binding motif, the third strand is all pyrimidines and binds to only one of the two strands of DNA by Hoogsteen bonding, wherein thymine recognizes adenine-thymine and protonated cytosine recognizes guanine-cytosine.

The targeting of DNA offers significant advantages over the antisense approach, in part because there are fewer target molecules per cell. For an average gene, approximately 1,000 copies of the corresponding mRNA may be present per cell. In addition, the RNA of a transcriptionally active gene is constantly being synthesized, thus requiring sustained concentrations of the antisense inhibitor. By contrast, duplex DNA — the target for triple helix agents — is present at only a few copies per cell in the case of endogenous genes or latent viruses, and its regeneration rate is typically far less than that of RNA. Thus, triple helix agents offer the prospect of low-dose, long-acting therapeutics.

The demonstration of sequence-specific alkylation triggered by triple helix binding gives rise to another advantage of the triple helix approach⁸. It is now conceptually possible to disable disease-causing genes by sequence-specific alkylation. As a result, ef-

fective treatments may be possible for cancers and viral diseases that involve transformed or persistently infected cells. For example, these advances open the exciting possibility of directly disarming the proviral genome of human immunodeficiency virus. By contrast, triple helix binding without alkylation should provide reversible inhibition and, like the antisense strategy, simply modulate protein synthesis in order to ameliorate disease, but at significantly lower doses.

The triple helix approach is limited, however, in terms of code recognition. For the Hoogsteen motif, only the purines, adenine and guanine, are recognized. In order to achieve sufficient specificity, a sequence of 15–18 purines must be identified in the target DNA. Such sequences do occur in clinically relevant genes, but only rarely. To overcome this limitation in addressable sequences, a "switchback" linker has been designed by molecular modelling to allow for the third strand oligonucleotide to bind with high affinity and specificity to both strands of the target duplex⁹ (see figure). This switchback between the two strands of DNA, which allows one segment of the oligonucleotide to recognize one strand of the double helix, and the second segment to read the other strand, greatly increases the number of sequences in the DNA that can be specifically targeted¹⁰. Also, by incorporating two or more alkylating moieties within a switchback oligonucleotide analogue, site-specific alkylation of both strands is possible, and the resulting complex is unlikely to be a substrate for DNA repair enzymes.

In summary, triple helix agents can bind in a sequence-specific manner to DNA, and thus potentially can be used to interrupt selectively gene expression. Triple helix agents have already been shown to block enzymatic cleavage of DNA and transcription factor binding in a site-specific manner¹¹. It remains to be seen whether site-specific triple helices can alter local DNA conformations so as to activate, rather than suppress, transcription.

Aptamers

Over the past few months, the opportunities for the development of oligonucleotides with therapeutic value have increased dramatically with the publication of techniques

Nucleotides and peptides

A new reagent for on-column biotinylation of oligonucleotides, a neurotoxic peptide antagonist at brain NMDA receptors and systems for multiple peptide synthesis are included in this week's review.

to generate compounds with high affinity to biological target molecules^{1,12-14}. These procedures involve the synthesis of random pools of oligonucleotides containing approximately 10^{13} different molecular species, each having a different nucleotide sequence. These pools are then incubated with the target molecules, and substances that bind with the highest affinity are isolated by physical separation techniques such as affinity chromatography or filter binding. The next step involves amplification of the isolated pool by enzymatic procedures; then the binding, selection and amplification cycles are repeated until the pool is enriched with only those oligonucleotides that have the greatest affinity. This technique allows for the selection of oligonucleotides that, by chance, have the correct three-dimensional structure to bind to the target molecule. In subsequent steps, the high-affinity oligonucleotides are evaluated for their ability to inhibit the activity (for example, enzymatic activity) of the target to which they bind.

Aptamers can be selected for affinity to large proteins¹²⁻¹⁴ and small organic target structures¹. Thus, high-affinity inhibitors potentially can be discovered for any extracellular target molecule for which a therapeutic benefit may be derived. The targeting of extracellular molecules circumvents the problem of permeation that must be faced if antisense oligonucleotides and triple helix agents of therapeutic value are to be developed. Notably, aptamer selection steps can be manipulated so as to subject the aptamer pool to more criteria than mere affinity for a given target molecule. Thus, other properties that are essential for therapeutic success can be conferred upon the final oligonucleotide.

In summary, oligonucleotides are a versatile class of molecules that have a varied and expanding range of biological activities. Fortunately, the chemical and biological tools are in hand to perform the difficult task of translating these biological activities into human therapeutics. □

Michael L. Riordan and John C. Martin are at Glend Sciences, 346 Lakeside Drive, Foster City, California 94404, USA.

DMT-biotin- C_6 -PA is the new **biotin phosphoramidite reagent** from Cambridge Research Biochemicals, which can be used for the on-column 5' biotinylation of oligonucleotides (*Reader Service No. 101*). The biotin phosphoramidite permits the synthesis of biotinylated oligonucleotides in much higher yields than with conventional off-column procedures, and in less time, says CRB. The reagent is soluble in acetonitrile and can be used with normal synthesis protocols. CRB claims coupling yields of greater than 95 per cent are possible, which, in many applications, allows the biotinylated oligonucleotides to be used without the need for purification. A special feature of the reagent is the dimethoxytrityl (DMT) protecting group on the biotin, which accounts for its solubility in acetonitrile. Removal of the DMT group by standard on-column detritylation provides users with a method of assessing the coupling yield. Biotinylated oligonucleotides can be used in dot blots, Southern blots, *in situ* hybridization and the polymerase chain reaction.

The Micro-Gel₁₀₀ **oligonucleotide analysis kit** from Applied Biosystems is intended to provide users with a rapid check of oligonucleotide purity (*Reader Service No. 102*). The kit, which is designed for use with the company's capillary electrophoresis systems, includes gel-filled capillaries, standards and the appropriate buffer. Micro-Gel₁₀₀ capillary electrophoresis is intended to offer advantages over other methods of purity analysis, such as slab gels and HPLC. The method is quick: it takes less than ten minutes to determine the purity of a 23-base oligonucleotide, says Applied Biosystems.



Check for oligo purity by capillary electrophoresis with the Micro-Gel₁₀₀ kit.

The high-throughput capability of capillary gel electrophoresis means that researchers need no longer use unqualified synthetic oligonucleotides. Micro-Gel₁₀₀ electrophoresis can be used to separate oligonucleotides up to 120 bases long with single-base resolution, whereas HPLC can only separate oligonucleotides of up to 30-40 bases with single-base resolution, says the company.

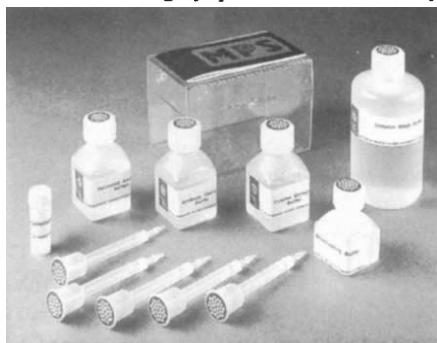
Antisense oligonucleotides are known to be sensitive to degradation by cellular DNases. In an attempt to tackle this problem, Biometra has developed oligonucleotides using **phosphorothioate derivatives of the parent nucleotides**, or PTOs (*Reader Service No. 103*). PTOs are highly resistant to cellular DNase attack but, unlike all other oligonucleotides, are still able to form a suitable substrate for RNase-H when hybridized with mRNA, says Biometra. PTOs against various proto-oncogene mRNAs and growth factor mRNAs are available, all of which are tested to inhibit specifically translation of their target mRNAs in cell cultures and tissue preparations. All PTOs can be supplied with a control PTO of the same base composition but randomized sequence of the relevant antisense oligonucleotide. The company will also custom synthesize antisense PTOs complementary to any desired sequence.

Peptide paraphernalia

Conantokin-G, a **neurotoxic peptide antagonist** at brain NMDA receptors, has been added to Bachem's armoury of neurotoxins (*Reader Service No. 104*). The 17-amino acid synthetic peptide Gly-Glu-Gla-Gla-Leu-Gln-Gla-Asn-Gln-Gla-Leu-Ile-Arg-Gla-Lys-Ser-Asn-NH₂ has no disulphide bonds and five residues of the unusual amino acid, gamma-carboxyglutamate (Gla). The "Gla" residues in conantokin-G chelate Ca²⁺ ions; this provides stability by promoting alpha-helix formation, says Bachem. It has also been proposed that binding of conantokin-G to receptors on acidic cell membranes is facilitated by Gla residues. Bachem says that recent experiments indicate that conantokin-G and -T are antagonists of brain NMDA receptors, a subtype of glutamate receptors. It has been suggested that induction of sleep in mice that are less than two-weeks-old, but hyperactive in mice that are older than three weeks, is a function of NMDA receptor blockage by these peptides. Conantokin-G can be purchased in 0.1-, 0.5- and 1.0-mg quantities for \$150 (US), \$600 (US) and \$1,000 (US), respectively.

1. Ellington, A.D. & Szostak, J. *Nature* **346**, 818-822 (1990).
2. Izant, J.G. & Weintraub, H. *Cell* **36** 1007-1015 (1984).
3. Melton, D.A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 144-148 (1985).
4. Uhlmann, E. & Peyman, A. *Chem. Rev.* **90**, 543-584 (1990).
5. Shaw, J.P., Kent, K., Bird, J., Fishback, J. & Froehler, . *Nucleic Acids Res.* **19**, 747-750 (1991).
6. Matteucci, M. *Tetrahedron Lett.* **31** 2385-2388 (1990).
7. Moser, H.E. & Dervan, P.B. *Science* **238** 645-650 (1987).
8. Povsic, T.J. & Dervan, P.B. *J. Am. chem. Soc.* **112**, 9428-9430 (1990).
9. McCurdy, S., Moulds, C. & Froehler, B. *Nucleosides Nucleotides* (in the press).
10. Horne, D.A. & Dervan, P.B. *J. Am. chem. Soc.* **112**, 2435-2437 (1990).
11. Maher, L.J., Wold, B. & Dervan, P.B. *Science* **245**, 725-730 (1989).
12. Tuerk, C. & Gold, L. *Science* **249**, 505-510 (1990).
13. Blackwell, T.K. & Weintraub, H. *Science* **250**, 1104-1110 (1990).
14. Blackwell, T.K., Kretaner, L., Blackwood, E.M., Eisenman, R.N. & Weintraub, H. *Science* **250**, 1149-1151 (1990).

Using the ProtOn Kits 1 and 2 from Multiple Peptide Systems, peptide users can not only **purify an anti-peptide antibody** but also affinity purify the native protein from which the peptide was derived (*Reader Service No. 105*). The purification procedures can be completed in 2.5 hours, says MPS; all the necessary columns, buffers and instructions are supplied with the kits. Using the ProtOn Kit 1, anti-peptide antibody is purified from immunized animal serum after immobilizing the peptide of interest on the activated chromatography resin. The starting point with Kit 2 is the purified anti-peptide antibody obtained with Kit 1. The purified polyclonal antibody is immobilized on activated chromatography resin, which then functions as a highly specific immunoaffinity



ProtOn kits from MPS purify anti-peptide antibodies and native protein.

chromatography resin for the native protein from which the original peptide sequence was derived. The kit can be used for the purification of larger proteins when the starting peptide sequence lies on the exterior surface of the native protein.

Clontech is offering a custom protein/peptide **microsequencing service** (*Reader Service No. 106*). Samples of purified protein/peptide in the low picomole range can now be sent to Clontech for sequencing. Using liquid-phase, covalent attachment sequencing methods, Clontech claims it can sequence most types of sample, including samples electroblotted onto Immobilon PVDF membranes. After receipt of the purified protein/peptide sample, the turnaround time for returning sequence data for 15–20 N-terminal residues is 2–3 weeks. The sequence is returned along with sequence documentation, which includes HPLC chromatograms of PTH derivatives with retention times, peak areas and height. Comparative HPLC plots of sequence cycle and standard PTH confirmations are also provided.

Peptide synthesis

The Novasyn Gem semiautomatic, solid-phase **peptide synthesizer** from Novabiochem combines the benefits of mild Fmoc polyacrylamide chemistry with the latest flow technology (*Reader Service No. 107*). The instrument has two reaction columns of varying size, which provide the user with the flexibility to scale up or to make two ana-



Novabiochem's peptide synthesizer has a two-column capability.

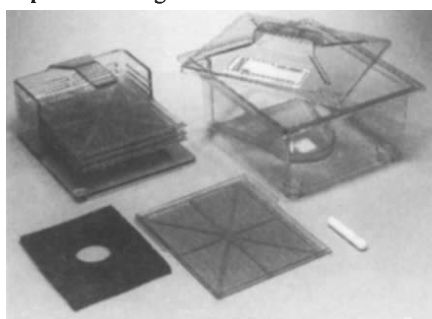
logues in the same synthesis. For most research purposes, the available scale of 0.05 mmol to 2 mmol is adequate for making milligram-to-gram quantities. With average cycle times of about an hour, 5–8 residues can be added to the peptide chain in a working day, says Novabiochem. There are no instrument-imposed limits on peptide length and the system is compatible with a range of activation chemistries. The Novasyn Gem incorporates a new monitoring system called Counterion Distribution Monitoring (CDM), which is designed to detect the rate of acylation while it is occurring. Consequently, interfacing the instrument with a spectrophotometer and chart recorder allows the system to take full advantage of acylation (CDM) and deprotection (UV) monitoring. This set up provides the assurance that all the steps have been successfully completed before the next amino acid addition, says the company.

In collaboration with scientists at the European Molecular Biology Laboratory, Abimed has developed a robotic system for **automated multiple peptide synthesis** called the AMS 422 (*Reader Service No. 108*). The instrument can be used for synthesizing 48 peptides simultaneously with an available scale that ranges from 5 μ mol to 50 μ mol, says Abimed. Syntheses are performed in individual column reactors that allow for washing of the resin in a flow-through mode; this arrangement is designed to eliminate the possibility of resin cross-contamination. Amino acid derivatives and activator are distributed by a metering syringe, whereas solvents are delivered via a manifold by nitrogen pressure. According to Abimed, total cycle times of less than two hours for 48 couplings can be achieved. Users can define synthesis parameters or, alternatively, can take advantage of the optimized routines that are designed to provide efficient peptide synthesis with minimal consumption of reagents and solvents. Operation of the robot and documentation of the synthesis parameters are controlled by a Macintosh computer.

BioExpo 1991

Paris, France will be the city hosting next week's BioExpo 1991 exhibition of biotechnology as applied to research, industry and agriculture. Some of the companies and the new products that will be featured at BioExpo are listed below.

The clear plastic construction of the DS-5000 **gel destainer** from Gibco BRL allows users to view gels during processing (*Reader Service No. 109*). A stir bar and magnetic stirring plate are used to circulate staining or destaining solutions across the entire surface of a gel. Dye is removed from the solution as it passes through the stain-absorbent char-



Gibco BRL's DS-5000 gel destainer accepts mini-gels, tube and slab gels.

coal foam that is located at the bottom of the tank. The DS-5000 is fitted with a multiposition tray rack, which allows simultaneous processing of gels of varying formats, including mini-gels, tube gels and slab gels. Rotating the tray rack 90° allows excess destaining solution to be drained back into the tank, says Gibco BRL. The tray rack holds up to four gel trays; tray adapters for mini-gels and tube gels can be purchased separately. See Gibco on stand E30.

Beckman, exhibiting on stand H27, has introduced four new **UltraSpherogel silica-based, size-exclusion HPLC columns** designed for the separation and purification of proteins and peptides (*Reader Service No. 110*). The UltraSpherogel-SEC columns use size-exclusion chromatography to separate samples according to molecular size and/or shape. The new line up includes the 7.5 $\times$ 300 mm Models 2000, 3000 and 4000, each of which costs \$595 (US), and the \$250 (US) 7.0 $\times$ 40 mm Precolumn. Each column is packed with a pure metal-free silica support of 5 μ m spherical particles with polyether bonding. Beckman recommends the use of the Model 2000, which features a 145 Å pore size and a molecular weight



Four new UltraSpherogel size-exclusion columns — see Beckman.

range of 1–250 kilodaltons, for the separation of small proteins and protein fragments. For medium-size proteins, the Model 3000 can be used, which has a 245 Å pore size and a range of 5–700 kilodaltons. The Model 4000 is suitable for separating large hydrophilic molecules and agglomerates. It features a 350 Å pore size and a molecular weight range from 10 to greater than 2,000 kilodaltons.

In response to an increasing demand for documentation in microscopy, Olympus has introduced three new **research microscopes** called the AH-3 Series (*Reader Service No. 111*). All models feature two permanently mounted 35-mm cameras, a large format capability of up to 10 inch × 8 inch and video port. In addition, all models have assisted focus, fully automatic real-time exposure control, automatic reciprocity correction at eight levels, one per cent spot metering, con-



stant illumination quality and perfect framing via the NFK turret and rotating stage mount. The AHBS-3 fully automated microscope for life science applications is able to focus automatically from any stage height and up to a 40× objective. Automation has also been applied to Koehler illumination and brightness stabilization, which are particularly important for video applications. Where full automation is not a requirement, Olympus recommends the AHBT-3 microscope. A triple module condenser features changeover and a motorized nosepiece makes the model particularly suitable for fluorescence microscopy, says Olympus. The AHBT-3 has a range of observation modes: brightfield, darkfield, phase-contrast (positive or negative), Nomarski DIC and polarized light. At 58 mm, Olympus claims that the AHMT-3 material science research microscope can accept the greatest specimen height in its class. See stand C15.

Cell culture aids

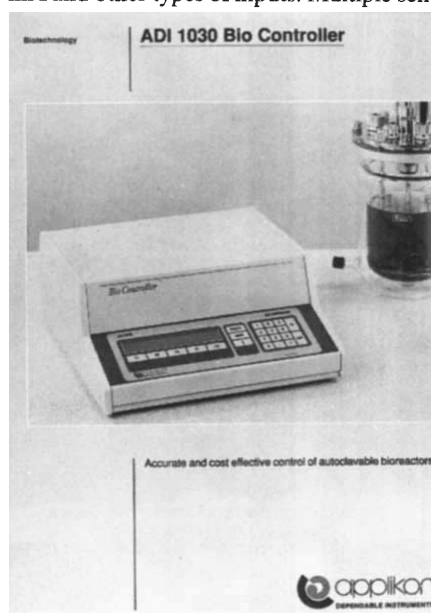
Mycoplasma contamination, the bugbear of cell culture laboratories, can be detected with the new **ELISA mycoplasma detection kit** from Boehringer (*Reader Service No.*

112). The test takes less than 24 hours from start to finish, providing a useful alternative to existing methods of detection that rely on either DNA staining or culture techniques. According to Boehringer, the kit can be used to detect contamination down to 10⁴ CFU per millilitre. The results are interpreted by visual examination of the microtitre plate for colour development. Consequently, no special expertise or equipment is necessary. The mycoplasma detection kit includes antibodies to the four most common mycoplasma/acholeplasma, plus the necessary buffers and substrates. Visit Boehringer on stand G17.

Imocell is the new **serum-free medium** from I.B.F. for the culture of hybridoma and lymphoblastoid cell lines (*Reader Service No. 113*). The medium contains all the necessary components for cell growth and high protein secretion, particularly the secretion of monoclonal antibodies, says I.B.F. Imocell is tested to ensure that it is free of mycoplasma and viruses and for batch-to-batch variability. It has been used successfully in a variety of cell culture systems, including macrowell plates, flasks, roller bottles and bioreactors. Imocell costs FFr230 (France) per litre. I.B.F. will be on stand E18.

Bioprocessing control

The new ADI 1030 **biocontroller** from Applikon — see stand E33 — is designed to provide accurate control over the conditions used to cultivate cells or microorganisms in autoclavable bioreactors (*Reader Service No. 114*). The instrument can control up to four parameters through one or more of the 11 outputs that are located at the rear of the instrument. Sensor cards are available for the control and measurement of pH, temperature, DO₂ (polarographic), DO₂ (galvanic), foam, CO₂, mV/redox, isolated mV/mA and other types of inputs. Multiple sen-



ADI 1030 biocontroller — see stand E33.

ADVERTISEMENTS

DNA FINGERPRINTING

Fresenius AG, Bad Homburg is offering the following oligonucleotides **exclusively**, which were evaluated and **patented** by Dr. J.T. Epplen and colleagues, Max-Planck-Institut für Psychiatrie, Martinsried, F.R.G:

(GTG)₅ · (GATA)₄ · (GACA)₄ (GGAT)₄ · (TCC)₅

These probes are available in unlabeled (10 pMol/aliquot) or Digoxigenin labeled (140 pMol/aliquot) form; lyophilized.

Fresenius

Diagnostic Division

6370 Oberursel, F.R.G. · Telephone: + 61 71/60-2443
Telefax: + 61 71/60-21 35 · Postbox 1809

Reader Service No.78

LIPOSOMES

REPRODUCIBLE — HOMOGENEOUS —
EXTREMELY RAPID



Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5–10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

DIANORM

POB 126, D-8000 Munich 65, FRG

Reader Service No.17

TRITIATED & ULTRA PURE PEPTIDES

An increasingly recognized problem is the purity of peptides used in biological studies. We offer synthesis of peptides including chemical synthesis of tritiated peptides of a very high purity.

SERTEK

Haraldsgade 68, DK-2100 Denmark
Tel: +45-3120 2554; Fax +45-3181 7833

Reader Service No.18

ADVERTISEMENTS

Severn Biotech Ltd.
Premixed Acrylamide Gel Solutions
 For COSHH safe PAGE the easy way with:
 Premixed Acrylamide Gel Solutions.
'Design a Gel'
BUY SEVERN BIOTECH
 For further product information
 Tel: 0562 825286 Fax: 0562 825284
 Unit 23, Hoobrook Enterprise Centre,
 Worcester Road, Kidderminster DY10 1WY

Professional Custom Services
DNA SYNTHESIS
PROTEIN SEQUENCING
PEPTIDE SYNTHESIS
 Rapid Service · Guaranteed Quality Ample
 Quantity
 Competitive Prices · Free Quotations
 Worldwide Delivery
 V.G. Services,
 1200 Wonderland Rd,
 S. Bldg. #7, Unit #2, London,
 Ontario Canada N6L 1A8
 Call/Fax: (519) 652-3758 Alt. Fax: (519) 652-9622

Custom DNA
 Purified and Shipped in 48 hours,
 Worldwide.
 100 ug. \$10.00 a base for the first 15 bases,
 \$7.50 for each additional base.
 Delivered.
Research Genetics
1-800-533-4363
 In the United Kingdom 0-800-89-1393
 In Canada 1-800-533-4363
FAX 205-536-9016

O₂/CO₂ RESPIROMETER
 measures bacterial and fungal growth



- * IBM-PC based.
- * Fully automatic 24hour measurements.
- * Periodic air refresh.
- * 1-20 measuring chambers (50ml-20 L).
- * Sensitivity 0.2µL/h.

For more information please contact:
Columbus Instruments
 P.O. Box 44049 Columbus, Ohio 43204 USA
 PH: (614) 488-6176; 1-800-669-5011
 TLX: 246514 FAX: (614) 276-0529

FASEB 1991 April 22-25, 1991
Booth Nos 1331 & 1333

sor inputs, multiple control outputs, bi-directional computer communications and menu-driven displays are provided.

Acudata is a new software package from Endotronics, which is designed for use with the company's Acusyst-R, Acusyst-Jr, and both the Acusyst Maximizer 500 and 1000 cell culture systems (Reader Service No. 115). Using Acudata to calculate metabolic rates and project perfusion rates, the process control of these instruments can be optimized, says Endotronics. By entering data obtained from metabolic and product assays, Acudata will calculate the glucose uptake rate, lactate production rate and product formation rate. The user can also select

ENDOTRONICS		Version: 1.13B
ACUDATA		Data Entry
Day	Pump Rate Projection	
Time	Desired Glucose Conc. (mg/dl).....	
Medium Rate (ml/hr)	Desired Lactate Conc. (mg/dl).....	
Factor Rate (ml/hr)	Length of Projection (hours).....	
Base Rate (ml/hr)		
Glucose Rate (ml/hr)	Recommended Medium Pump Rate (ml/hr) ..	140
Medium G/LC (mg/dl)	Recommended Glucose Pump Rate (ml/hr) ..	8
Factor G/LC (mg/dl)	Projected Glucose Conc. (mg/dl).....	255
Glucose G/LC (mg/dl)	Projected Lactate Conc. (mg/dl).....	99
Glucose (mg/dl)		
Lactate (mg/dl)		
G/LC (mg/hr)		
LPR (mg/hr)		
	APP-27 Cell Line	
	Page 2	
	File Name	

Cell culture optimization software.

glucose and lactate setpoints and Acudata will project the media and factor feed rates that are necessary to meet these setpoints. Run data can be displayed graphically for presentation and documentation purposes. Acudata software, available on 3.5- and 5.25-inch disks, runs on an IBM XT, AT, PS/2 or compatible computers. Cytolab, the French distributor for Endotronics will be exhibiting the company's cell culture systems on stand E34 at BioExpo.

Information sources

As of 21 March, the Immunoclonal Data base (ICDB) from CERDIC (Centre Européen de Recherches Documentaires sur les Immunoclonales) is available from Data-Star (Reader Service No. 116). The ICDB of CERDIC draws information on hybridomas and other cell lines of immunological interest from over 1,500 scientific journals, as well as commercial catalogues and international patent applications. Each record con-

tains both scientific data on the reactivities of the monoclonal antibody and at least one contact name and address where the cell or product can be obtained. Approximately 23,000 records are presently available; 5,000 were added in 1990. CERDIC is associated with the European Commission's BRIDGE programme, a two-year project on information networks, which involves six centres in Europe—United Kingdom, Italy, Belgium, the Netherlands, Germany and France. CERDIC is available online on DIMDI (Germany), Data-Star (Switzerland) and through Minitel (France). For the first three months following the launch, there is no connect hour charge on Data-Star for ICDB. Users can search the data base free-of-charge until June, paying only for the documents printed: SFr2.40 (Switzerland) for a long document, SFr1.20 (Switzerland) for a short document. See CERDIC on stand E12.

The new 200-page cell culture catalogue and handbook from Sera-Lab features a wide variety of cell culture products, including both liquid and dry powder media, sera and serum alternatives, and antibiotics and factors required for growth, transport and



Media by mail order.

attachment of cells (Reader Service No. 117). A technical information section provides a useful reference guide. Details of manufacturing and testing procedures are included, together with information of the use of the products. Formulation details, a glossary and index are also included. See Sera-Lab on stand E14.

The 1991/1992 catalogue from Biozyme contains a host of enzymes for applications in clinical diagnostics, immunodiagnostics and biotechnology (Reader Service No. 118). Alcohol dehydrogenase, alkaline phosphatase, cholesterol esterase, enterokinase, glucose oxidase, glycerol-3-phosphate dehydrogenase, peroxidase, ribonuclease, trypsin inhibitor and urease are a selection of the products listed. New product entries include ascorbate oxidase and many improved grades of enzymes. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain further details, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.

Synthetic Oligonucleotides

Custom made probes and primers,
 synthesised, purified and shipped in 3 days

£3.50 per base

(All inclusive)

0.2 µM scale (approx. 500 µg)

Modified bases, biotin or digoxigenin labelling,
 data base analysis and full design facilities

Molecular Medicine Unit,
 King's College School of Medicine & Dentistry,
 Denmark Hill, London SE5 8RX

Tel. XX-44-(0)71 - 326 3126

FAX XX-44-(0)71 - 733 3877

Reader Service No.56